Non-blind audio-watermarking using compression-expansion of signals

Foo, Say Wei

2008

Foo, S. W. (2008). Non-blind audio-watermarking using compression-expansion of signals. In Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems 2008: (pp.1288-1291). Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/91128>

<https://doi.org/10.1109/APCCAS.2008.4746263>

©2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. <http://www.ieee.org/portal/site>.

Downloaded on 26 Apr 2023 19:11:13 SGT

Non-blind Audio-watermarking using Compression-Expansion of Signals

Say Wei FOO
Nanyang Technological University
Singapore 639798

Abstract—One of the advantages of non-blind watermarking scheme is that the original signal must be available to extract the watermark. As such, the possession of the original signal becomes important in the proof of ownership. In this paper, a non-blind audio-watermarking method is proposed. The method makes use of compression-expansion of adjacent frames of samples of the audio signal. Some frames of the signal are thus ‘modified’. The modified frames can be readily identified by comparing them with the original signal. The modified frames can be used to embed hidden information. By applying psychoacoustic model in the compression-expansion process, it is found that the modification does not affect the quality of the audio signal.

The embedding of watermarks for single channel audio signal and two-channel stereo signals require different approaches which are described in detail in the paper. The performance of the method under several types of attacks is assessed. It is found that high percentage of recovery can be achieved with spectrally rich audio signals.

Keywords—Non-blind technique, audio, watermarking, compression, expansion.

I. INTRODUCTION

Audio-watermarking is used mainly to identify the owner of the audio signal although it can be used as a way to store hidden information. Ability to recover the watermark is more important than the efficiency of storage of information if the purpose is to identify ownership. Depending on how watermarks are embedded in the audio signals, the original signals may or may not be required in the recovery process. For non-blind watermarking methods, the original signals are required to compare with the watermarked signals in order to recover the watermarks. Such methods have the advantage that only the owner will have the original signals to prove his ownership. However, an inventory of the original signals must be kept and safeguarded.

In this paper, a non-blind audio watermarking technique is described. Since most audio signals make use of two-channels for stereo effect, the use of signals in both channels are also investigated in addition to watermark embedding using signal from one channel only. It is shown that the employment of two channels has increased the flexibility of watermark embedding. At the same time, it also increases the probability of watermark recovery especially when the signals are subjected to further signal processing or so-called ‘attack’.

The organization of the paper is follows. The basic idea of the technique is described in Section II. Watermark

embedding methods using single channel signals and stereo signals are detailed in Section III. The performance of the methods is presented in Section IV; followed by concluding remarks in the last section.

II. THE WATERMARK BIT

A. Representing Data in Binary System

One way to represent watermark is to use a string of letters of the English alphabet. If we restrict the letters to the upper case, we only need 26 symbols and these can be coded using 5 bits ($2^5=32$).

To make use of the binary system to represent information, there must be two distinguishable states. For the proposed watermarking system, we make use of compression-expansion of consecutive frames as one state, which we shall refer to as watermark bit. When the compression-expansion frames are subtracted from the original signal frames, the difference waveform resembles the shape of a diamond. We shall also refer to this watermark bit as diamond. Details of the diamonds are given in the following section.

B. The Diamonds

As in most frame-based one-dimensional process, the audio signal is segmented into frames with 50% overlap. Two consecutive frames that are suitable for embedding of watermark bit are identified. Hanning windows are then applied to the frames. The windowed signals are then transformed into frequency domain using Discrete Cosine Transform (DCT). The signals in both domains will have equal number of samples.

The watermark bit is created as follows. For the first of the two frames, some (4 out of a frame of 512 samples) transform coefficients are removed, after which inverse DCT is applied to recover the time-domain signal. The frame of the signal is thus compressed. In order not to affect the overall length of the audio signal, the same number of transform coefficients is added to the DCT of the second frame and inverse DCT applied to recover the time-domain signal. The second frame of the signal is thus expanded. The duration of the two frames taken together, however, remains the same as that of the original signal. The time-domain waveform of these two frames that are subjected to compression-expansion is thus modified. This distortion is used to represent one watermark

bit. By using psychoacoustic model [11,12] in the decision to remove or insert transform coefficients, the perceptual quality of the resulting audio signal should not be significantly affected.

The embedding of the watermark bit is thus by compression-expansion of consecutive frames of the signal. As the frames are thus ‘modified’, the modified frames can be detected by comparing them with the waveform of the original frames. By subtracting the waveform of the ‘modified’ signal from the waveform of the original signal, a difference signal whose envelop looks like a diamond as shown in Fig.1 is obtained. Note that the length of one ‘diamond’ is at most $3/2$ times the window/frame length, which is the total length of two consecutive frames with 50% overlap, while the shape of the diamonds depends on the amount of resulting distortion.

Figure 1. A typical diamond.

C. Effect of Reversing Compression Expansion

Having discussed the method of generating one watermark bit (say binary 1) by compression-expansion of consecutive frames, one might jump into the conclusion that by reversing the order of compression-expansion, i.e., expansion first followed by compression, one might obtain another watermark bit (say binary 0). However, experiments show that this is not the case. While the shapes of the diamonds are different using different order of compression-expansion of the same two frames, there are no readily distinguishable differences such that they can be used to represent different pieces of information such as binary ‘1’ and ‘0’. Thus, another way must be found to represent the other watermark bit. This will be discussed separately under the sections on single channel embedding and two-channel embedding.

It is noted that since there is a 50% overlap of frames, to avoid inter-symbol interference, one overlap frame after the two selected consecutive frames should be left out for consideration in the embedding process. In other words, each pair of modification should be separated by at least an untouched frame as illustrated in Fig.2.

It is apparent that frames with low energy contents may not be badly modified after compression-expansion and hence may not produce diamond that can be readily distinguished; and frames such as silent frames produce no diamond at all. As such, only those frames with energy level exceeding a given threshold should be selected.

Figure 2. Frame selection.

III. WATERMARK EMBEDDING AND RECOVERY

A. Embedding Watermark in Single Channel Signals

To make use of the binary system to represent information, there must be two distinguishable states. The compression-expansion of frames, the diamonds, can be taken as one state, say binary 1. The non-changed frames may be used to represent ‘0’. To use this method, a number of consecutive frames must be found so that all the required bits can be accommodated consecutively. For example, if we are to use BAT as a watermark, and each letter is to be coded with 5 binary bits, then we need $15 \times 3 = 45$ consecutive frames in addition to header frames. Note that three frames are required to encode one bit of information, At least one ‘diamond’ must be used as a header to signal the start of the bit sequence. Subsequently, if the frames are modified, a ‘diamond’ is obtained after decoding, then the symbol is ‘1’, if they are not modified, then the coded symbol is taken as ‘0’.

Figure 3. Watermark embedding

As the non-modified frames are used to represent the other binary bit, so the bits must be arranged strictly one after another so that the ‘0’ state can be identified. This is a

constraint of embedding using single channel signals. For certain audio signals such as speech signal, it may not be easy to find 45 consecutive frames that have sufficient energy in all the frames. If one of the frames does not have sufficient energy as shown as the last diamond in Fig.3(a), then recovery of ‘diamond’ and hence the watermark may be affected.

B. Embedding Watermark in Two Channel Signals

For stereo music, two channel signals are available. It is proposed to make full use of the two signals to do away with the constraint on the embedding process and to increase the probability of watermark recovery.

For a ‘diamond’ or binary 1 embedded in one channel, we will have a corresponding ‘null’ of binary 0 embedded in the other channel and vice versa as shown in Figs.3(b) and 3(c). In this case, the binary bits or diamonds need not be placed consecutively. It is then not necessary to find a large number of consecutive frames for watermark embedding. Only when two consecutive frames have sufficient energy, they will be used to code a binary bit in one channel; the corresponding ‘null’ in the other channel serve to confirm the location of the watermark bit. If the corresponding frames in both channels are ‘null’, then it is not taken as an information bit.

A header sequence such as five consecutive binary 1’s can be used to identify the start of the watermark bit sequence.

This two-channel approach to embed watermark data bits does away with the requirement to find for example, 45 consecutive frames for encoding. The diamond frames can be located at different parts of the signal where the signal strength allows easy detection.

B. Methods of Watermark Recovery

Recovery of the watermark is essentially the detection of the presence or absence of ‘diamonds’ or the modified frames.

One way is to detect the slope of the diamond envelop [1]; another way is to compute the energy contents of the relevant frames. For the two-channel watermarking method, we shall use the moving average root-mean-square value of the diamond signals. For audio signals sampled at 44100 samples per second, the averaging is carried out over 200 samples.

For the two-channel embedding approach, a presence of watermark in one channel must be complemented by a ‘null’ in the other channel to be considered as an information bit. Thus, in addition to the energy contents of one channel, the relative values of the energy contents of the two channels for the same time interval are also used to check the presence or absence of ‘diamonds’.

After the bits are determined, they are grouped in 5-bits and the symbol represented by the 5-bits is decoded. Some synchronization and shift matching may be incorporated if necessary. The diamond signals obtained when white noise is added to the watermarked signal are shown in Figs.4(a) and (c). The corresponding moving average rms values of the signals are shown in Figs.4(b) and (d). It can be seen that by

comparing the values in the two channels, the addition of noise does not impose a problem in watermark recovery.

Figure 4. Watermark recovery using moving average of root-mean-square values

IV. PERFORMANCE OF THE METHODS

The methods are assessed using several types of audio signals. The signature consisting of three letters of the English alphabet as represented by 3x5 bits codes are embedded in the signals. The methods of embedding and watermark recovery are as described in Section III. The following five audio pieces are used: 1) Classical Music, 2) Instrumental Music, 3) Rock Music, 4) Pop Music and 5) Male Speech. Rock Music and Pop Music have very high signal energy while Classical Music and Instrumental Music have moderate signal energy. The speech clip has a relatively high percentage of silent intervals.

The audio quality of the watermarked signals is evaluated through listening tests. The listeners were presented with the original signals and the watermarked signals in random and asked to indicate their preference. It is found that the embedded watermarks do not affect the audio quality to any significant degree.

Recovery is perfect for all audio signals tested when they are not attacked by further signal processing.

The watermarked signal is then put through various additional signal processing or so called attacks to test its robustness.

The following five types of attack (further signal processing) are applied on the watermarked signals.

- 1) Addition of White Gaussian Noise: White Gaussian noise is added so that the resulting signal has a Signal-to-Noise Ratio of 30dB.

- 2) Resampling: The test signal originally sampled at 44.1 ksamples/s is re-sampled at 22.05 ksamples/s, and then restored by sampling again at 44.1 ksamples/s.
- 3) Cropping: Some segments of the watermarked signal are replaced with segments of the signal attacked with filtering and additive noise.
- 4) Low-pass Filtering: A Finite Impulse Response (FIR) low pass filter with cutoff frequency at 10 kHz, passband ripple 0.173dB and stopband attenuation 40dB is used.
- 5) MP3 Conversion: The watermarked test signal is encoded into a 128kb/s MP3 format and then restored.

The rates of recovery of signature for both single channel and two-channel embedding are tabulated in Table I.

TABLE I. RATES OF RECOVERY OF SIGNATURE

Types of Attack	Types of Signal	Single Channel (%)	Two-Channel (%)
Addition of noise	<i>Classical</i>	100	100
	<i>Instrumental</i>	89	100
	<i>Rock</i>	100	100
	<i>Pop</i>	90	100
	<i>Speech</i>	100	100
Re-sampling	<i>Classical</i>	90	100
	<i>Instrumental</i>	100	100
	<i>Rock</i>	78	97
	<i>Pop</i>	22	89
	<i>Speech</i>	80	97
Cropping	<i>Classical</i>	100	100
	<i>Instrumental</i>	100	100
	<i>Rock</i>	100	100
	<i>Pop</i>	100	100
	<i>Speech</i>	60	100
Low-pass filtering	<i>Classical</i>	90	97
	<i>Instrumental</i>	89	97
	<i>Rock</i>	89	97
	<i>Pop</i>	22	93
	<i>Speech</i>	80	93
MP3 compression	<i>Classical</i>	100	100
	<i>Instrumental</i>	97	100
	<i>Rock</i>	11	78
	<i>Pop</i>	11	78
	<i>Speech</i>	60	80

From the results, it can be seen that the two-channel approach gives significantly better performance than the single channel embedding method. This is because the ‘diamond’ frames can be selected at suitable places of the audio signal and also by comparing the energy contents of the two channel

for the same time interval, the precision of detection of watermarks is increased.

V. CONCLUSION

A compression-expansion type of audio watermarking scheme is presented. Embedding the watermark bit in single signal requires a number of consecutive frames of signal with high energy contents. By using two signals, the same type of modification of the signal, in this case, compression-expansion of consecutive frames of the audio signal can be used to represent two different binary states in the two channels. Long string of frames is thus not necessary using the two-channel approach. In addition, by comparing the signals in the two-channel, more accurate recovery of watermark bits can be achieved.

Experimental results show that the performance of the two-channel approach in detection of watermark signature is significantly better than that using a single channel alone.

REFERENCES

- [1] Foo, Say Wei, Ho Shuet Mun, and Ng Ling Mei "Audio watermarking using time-frequency compression expansion" IEEE International Symposium on Circuits and Systems, 2004, pp. 201-204.
- [2] Foo, Say Wei, Xue Feng, and Li Mengyuan "A blind audio watermarking scheme using peak point extraction", IEEE International Symposium on Circuits and Systems, 2005, pp. 4409-4412.
- [3] Foo, Say Wei, Yeo Theng Hee, and Huang Dong Yan, "An adaptive audio watermarking system", IEEE Tencon 2001, pp. 509-513.
- [4] Acevedo, A.G., 2005. Audio watermarking: properties, techniques and evaluation. In Lu, CS. (ed.) *Multimedia security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*, Idea Group Publishing. Pennsylvania.
- [5] Cvejic, N. and Seppänen, T., 2005. Audio watermarking: requirement, algorithms, and benchmarking. In Seits, J, *Digital watermarking for digital media*, Information Science Publishing, Pennsylvania.
- [6] Foote, J., Adcock J., and Girgensohn, A., 2003. Time base modulation: a new approach to watermarking audio [Electronic Version], Palo Alto Laboratory. California.
- [7] Suhail, M.A., 2005. Digital watermarking for protection of intellectual property. In Lu, C.S, *Multimedia security: steganography and digital watermarking techniques for protection of intellectual property*, Idea Group Publishing. Pennsylvania.
- [8] Bassia, P, Pitas, IP, "Robust Audio Watermarking In Time Domain", EUSIPCO 1998, 8-11 Sept., Patras, Greece, pp. 25-28.
- [9] D. Gruhl, A. Lu, W. Bender, "Echo Hiding", in Proc. Information Hiding Workshop, University of Cambridge, U.K., 1996, pp. 295-315.
- [10] Seits, J., and Jahnke, T., 2005. Digital watermarking: an introduction. In Seits, J, *Digital watermarking for digital media*, Information Science Publishing. Pennsylvania.
- [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] Yost, W.A., 2000. Fundamentals of Hearing – An Introduction. Academic Press.