Two-channel time-frequency audio watermarking

Hertanto, Richard Nathaniel; Foo, Say Wei

2007

Hertanto, R. N., & Foo, S. W. (2007). Two-channel time-frequency audio watermarking. In Proceedings of the International Conference on Information, Communications & Signal Processing 2007: Singapore: IEEE.

<https://hdl.handle.net/10356/91127>

<https://doi.org/10.1109/ICICS.2007.4449719>

International Conference on Information, Communications & Signal Processing © 2007 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 15:19:16 SGT

Two-Channel Time-Frequency Audio Watermarking

Richard Nathaniel Hertanto
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore

Say-Wei Foo
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore

Abstract—The time-frequency audio watermarking scheme [1] operates by removing and adding samples to the frequency spectrum of the audio signals (compression-expansion) as a way to embed watermarks. The watermarks so embedded are identified by comparing the waveforms of the original and the watermarked signals. For a single channel signal, the compression-expansion represents one binary symbol and the unchanged frames in between are taken as the other binary symbol. For audio signals with two channels such as stereo music, a method is proposed in this paper so that compression-expansion in the two channels are used to represent different binary symbols. In addition, some modifications to the original scheme [1] are also presented.

Keywords—Audio, watermarking, time-domain, frequency-domain, two-channel.

I. INTRODUCTION

Digital watermarking can be regarded as the process of permanently embedding information into a digital data (audio, images and video) in such a way that the information can be later extracted using a computerized process/program. There are several applications for watermarking. These include: copyright protection, copy protection, copy prevention, content authentication, fingerprinting, broadcast monitoring, and data hiding.

There are many ways to embed watermarks in the audio signals [1-11]. Some methods are blind in that the original signal is not required for watermark extraction while others are non-blind in that the original signal must be used to detect the watermarks. Both types of techniques have their applications. For example, only the owner of the audio signal has the original signal, and only he could use it to extract the correct watermarks coded with non-blind technique to prove his ownership. The advantage of the blind watermarking schemes is that there is no need to safe-keep a large inventory of original signals.

In this paper, a non-blind audio watermarking technique especially for two-channel signals based on time-frequency conversion is discussed. The basic idea of the technique is given in Section 2, followed by methods used to test the robustness of the scheme against ‘attacks’ in Section 3. The results of experiments are given in Section 4. Concluding remarks are presented in the last section.

II. THE PROPOSED SCHEME

A. The Diamonds

The audio signal is segmented into frames with 50% overlap. The energy of the first frame is computed. If the energy of the frame exceeds a given threshold, it is considered useable for watermark. The next frame is then similarly processed. When two consecutive useable frames are found, then one binary bit of the watermark is set as follows. Hanning windows are applied to both the frames. The windowed signals are then transformed into frequency domain using Discrete Cosine Transform (DCT). The signals in both time and frequency domains have equal number of samples. Transforming the signal into frequency domain has the advantage of de-correlating the signal and removing some values that will not change the shape of the waveform significantly.

To compress a frame of the signal, a small number of samples at the high frequency end of the DCT are removed. To ensure that the total length of the audio signal remains the same as the original, the next frame is expanded by adding the same number of samples for compression to the high frequency end of DCT. If the number of samples added/deleted is small compared with the window size, the perceptual effect is hardly noticeable. Alternatively, psychoacoustic model [11,12] can be adopted to determine the samples to be removed or added to minimize the loss of quality of the audio signal.

To identify the modification, the waveform of the compress-expand signal is subtracted from the waveform of the original signal. A difference signal whose envelop looks like a diamond is obtained. It is noted that the length of the diamonds is $3/2$ times the window/frame length, which is the total length of 2 consecutive frames with 50% overlap, while the heights of the diamonds are of variable heights depending on the original audio.

Fig.1 and Fig.2 depict the difference signals for two different sound samples. It can be seen that the envelopes of the difference signals resemble the shape of a diamond. For ease of subsequent description, we shall refer to such difference signal as diamond.

Figure 1: Diamond A.

Figure 2: Diamond B.

B. Effect of Reversing Compression Expansion

In the method explained above, a frame of the signal is compressed and the following frame is expanded. To check the effect of the reverse modification, the same consecutive frames are first compressed and then expanded. By separately modifying the same audio segments using both order, subtracting the result with the original, and finally making comparison between the 2 difference signals, it is found that while they are not identical, there are no distinguishing differences such that they can be used to represent different information such as binary '1' and '0'.

Either compression-expansion or expansion-compression can be used to indicate one state. For the method used in our experiments, the frames are first compressed and then expanded.

It is noted that inter-symbol interference arises if four consecutive frames are used to embed two watermarks. That is, the four frames are compressed, expanded, compressed and expanded in that order. To avoid the problem of inter-symbol interference, each pair of modification should be separated by at least an untouched frame. This untouched frame also serves as separator for watermark symbols.

Another obvious problem is that frames with low energy contents may not produce diamond that can be readily distinguished; and frames with negligible signal (silent frames) naturally produce no diamond at all. As such, only

those frames with energy level exceeding a given threshold are considered.

C. 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 be upper case, we only need 26 symbols and these can be coded using 5 bits ($2^5=32$). A watermark code may consist of three letters, for example: NTU and $3 \times 5 = 15$ bits are required to code this watermark.

To make use of the binary system to represent information, there must be two distinguishable states. Since compression-expansion and expansion-compression produce similar difference signals or diamonds, such modification can be used to represent only one state. One way of solving this problem is to use non-changed frames to represent '0' and diamond frames as '1'. To use this method, a number of consecutive frames must be found (for watermark NTU, 15 frames are required) that can be used for watermark embedding and header be used to signal the state of the 'bit sequence'. The diamond frame must be used as a header. Subsequently, three frames are considered at a time, if they constitute a diamond frame, then the symbol is '1', if they are not modified, then the coded symbol is taken as '0'. This is the approach adopted in [1].

D. Two Channel Binary Representation

As most of today's audio data are stereo encoded and so have two channels, it is proposed to use the diamonds in the two channels differently as described in detail in the following.

A diamond in one channel and a diamond in the other channel at the same time frames will be used to indicate the start of a watermark code.

Let the two channels be called Channel 1 and Channel 2. A diamond in Channel 1 and non-modification in the other channel, Channel 2 will be used to indicate the binary '1'.

A diamond in Channel 2 and non-modification in Channel 1 will be used to indicate the binary '0'.

Non-modification frames in both channels will not be considered as code.

This two-channel approach to encode watermark data bits does away with the requirement to find 15 frames together for the encoding. The diamond frames can be located at different parts of the signal where the signal strength allows easy detection and minimizes degradation.

E. Method of Decoding

The most significant feature of a modified section is the "diamond" shape envelop obtained after subtracting the

modified audio signal with the original signal. The decoding works by accurately detecting the presence or absence of the diamond frames. The diamond envelop can be isolated by taking the absolute value of the difference and then performing a low-pass filter operation. However, this approach by itself does not give good results especially with watermarked signals that are put through further processing (attack).

A novel approach to detect the diamond frames is described in the following. A reference diamond is first created using the shapes generated from several audio signals. The diamond envelops generated from different audio signals are first obtained. These various diamonds are normalized by dividing the values by the largest value. An average diamond shape is then obtained by taking the average of these normalized diamonds. A typical averaged diamond is shown in Fig.3.

Figure 3: Reference diamond.

Varying values are obtained by cross-correlating the reference diamond with envelop of diamond frames. Observation indicated that the range of values of cross-correlation of the reference diamond with that of modified frames gives values that range from 0.75 to 1 while the values for those with unmodified frames generally fall below 0.5.

In total, there are 3 ranges of values:

1. 0.75 to 1: decide as modified
2. 0.5 to 0.75: decide as probably modified
3. below 0.5: decide as not modified

Note that, as the encoding process skips frames that are too quiet as they are unsuitable for encoding, the decoding process must also perform this check on the original data and skip the frames accordingly. This will ensure that only frames suitable for marking are checked for processing.

As two channels are used for the watermarks, each channel has 3 possible outcomes, so there are total of 9 different combinations of cross-correlation values. Let the combinations be represented as $\{c1, c2\}$ where the two numbers in the braces being the cross-correlation values of Channel 1 and Channel 2 respectively. The combinations $\{1,1\}$, $\{1,2\}$ and $\{2,1\}$ may be taken as diamond frames in both channels and hence regarded as synchronization signal; $\{1,3\}$ may be taken as '1' and $\{3,1\}$

taken as '0' while $\{3,2\}$, $\{2,3\}$, $\{3,3\}$ may be taken as non-modified frames. A $\{2, 2\}$ may be regarded as ambiguous bit.

After the bits are determined, they are grouped in 5-bits and the symbol represented by the 5-bits is decoded.

III. TESTS OF ROBUSTNESS

The proposed audio watermarking scheme is put through various additional signal processing or so called attacks to test its robustness.

Three different categories of attacks are tested. They are:

1. Addition of white Gaussian noise such that the resulting Signal-to-Noise Ratio (SNR) is 30 dB.
2. Resampling: The signal is re-sampled at half the original sampling rate and again at the original sampling rate.
3. Compression: The watermarked signal in wav format is converted to one of the following formats: to MP3, OGG, and WMA all at 128 Kbps. It is then converted back to wav format for watermark detection.

IV. EXPERIMENTS AND RESULTS

Experiments are carried out to assess the performance of the proposed approach. Five types of audio signals: classical, instrumental, rock, pop music and speech signals are tested. The performance is measured by the letters correctly decoded by the scheme. The letters of the English alphabet as represented by 5 bits codes are embedded in the two channels. The modification carried out is compression followed by expansion. In other words, four samples of the identified frame are deleted followed by addition of four samples in the next frame. In the experiments, the last four samples of the DCT transform are deleted for compression and four samples are added to the DCT transform for expansion.

A diamond frame, or compression-expansion in Channel 1 together with non-modification in Channel 2 at the same time frame is treated as the binary bit '1'. A diamond in Channel 2 together with non-modification in Channel 1 is treated as binary bit '0'.

The watermarked signal is then put through various attacks mentioned in Section 3. After the attacks, the difference between the watermarked signal and the original signal is obtained for the two channels. The diamonds in the two channels are detected using the cross-correlation method mentioned in Section 2. After which the bits are grouped into 5 per group and the symbol it represents is identified.

The outcome of detection is put under one of the following three categories:

1. Success: the letter is correctly identified
2. Error: the wrong letter is identified.

3. Ambiguous: there might be an error in the bit.

The average detection rates for all five types of audio signals tested are tabulated in Table I. The best performance is obtained using classical music; the results are presented in Table II.

TABLE I. AVERAGE DETECTION RATES

	Success	Ambiguous	Error
No Attack	100%	0.00%	0.00%
Addition of Noise	91.30	8.70%	0.00%
Resampling	78.99	21.01%	0.00%
MP3	55.07	44.20%	0.72%
OGG	66.67	33.33%	0.00%
WMA	83.33	16.43%	0.10%

TABLE II. DETECTION RATES FOR A CLASSICAL PIECE

	Success	Ambiguous	Errors
No Attack	100.00%	0.00%	0.0%
Addition of Noise	100.00%	0.00%	0.0%
Resampling	85.71%	14.29%	0.0%
MP3	92.86%	32.14%	0.0%
OGG	100.00%	0.00%	0.0%
WMA	92.86%	7.14%	0.0%

From the results given in Table I, it can be observed that perfect recovery is possible with no degradation of the watermarked signal. The addition of noise has the least effect on the recovery of the letters and MP3 compression has reduced the accuracy of detection to almost half. However, the performance for different audio signals is different as observed from values given in Table II.

V. CONCLUSION

A two-channel time-frequency audio watermarking scheme is presented. By using two channels, 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 the two different binary states in the two channels.

The performance of the scheme in detection of letters depends on the types of attacks and the types of audio signals. Perfect recovery of the watermarks is assured for all types of audio signals if there is no attack on the watermarked signals. The scheme is least vulnerable to addition of noise but performs worst when subjected to MP3 coding and decoding.

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, C.S. (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 Girsensohn, 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.