Audio watermarking using time frequency compression expansion

Foo, Say Wei; Ho, Shuet Mun; Ng, Ling Mei

2004

Foo, S. W., Ho, S. M., & Ng, L. M. (2004). Audio watermarking using time frequency compression expansion. International Symposium on Circuits and Systems (pp. 201-204) 2004:Vancouver.

<https://hdl.handle.net/10356/90835>

<https://doi.org/10.1109/TSP.2007.893965>

©2004 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. 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>

AUDIO WATERMARKING USING TIME-FREQUENCY COMPRESSION EXPANSION

Foo Say Wei, Ho Shuet Mun and Ng Ling Mei

Dept of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798

ABSTRACT

A novel non-blind digital audio watermarking scheme using time-frequency based operations is presented. The digital audio samples are first segmented into frames and the frames are transformed into the frequency domain. Frequency components are then added or removed. The amplitude of the components added or removed are determined under the guidance of Psychoacoustic Model I. Inverse transform is then performed to restore it back to a time domain signal which is a lengthened (expanded) or shortened (compressed) version of the original. The expanded or compressed frames are then used to code the watermark to be embedded. Informal listening reveals that there is no perceptible distortion of the audio signal for suitably compressed or expanded segments. Recovery of watermark is perfect for signals not subjected to attacks. Experimental results also show a high degree of successful recovery of the signature that allows conclusive authentication of the owner even after filtering, cropping, MPEG compression, resampling and the addition of noise.

Keywords

Watermarking, data hiding, digital audio signal, time-frequency technique.

1. INTRODUCTION

With the availability of inexpensive Compact Disc (CD) and Digital Versatile Disc (DVD) writers and also the easy access to Internet, the storage and transmission of data have become easier and cheaper. This has contributed to the problem of unauthorized copying and distribution, making the protection of intellectual property an urgent issue.

Protection systems employing encryption, usage control and watermarking have been developed to protect intellectual property. Encryption and usage control entail inconvenience to the legitimate end users. Watermarking,

on the other hand, is a less obtrusive method of protection. While watermarking would not prevent unauthorized copying or distribution, it does provide the signal with a unique identity such that the owner can be traced.

In this paper, a new method of digital audio watermarking based on time-frequency domain operations is proposed.

The organization of the paper is as follows. Following this introductory section is a section giving an overview of the proposed scheme. A brief description of the Psychoacoustic Model is given in Section 3. Details of the watermark embedding process and the watermark recovery process are covered in Sections 4 and 5. The performance of the scheme is presented in Section 6 and the concluding remark in Section 7.

2. THE PROPOSED METHOD

The proposed audio watermarking scheme makes use of consecutive expanded (lengthened) and compressed (shortened) frames of samples to indicate the binary code of '1', and a normal frame to indicate the binary code of '0'.

A sequence of letters of the alphabet, such as *FSW* is used as the watermark. To convert the letters to binary format, the ASCII code is adopted. Thus, *FSW* is represented in binary form as 010001100101001101010111. This binary signature is then embedded in the signal as a watermark.

To minimize the perceptual distortion of the coded audio signal, Psychoacoustic Model is used to guide the processes of expansion and compression. A brief description of the Psychoacoustic Model is given in Section 3.

For the recovery of the signature from the watermarked signal, the original signal is matched with the watermarked signal and the difference signal is obtained. From the difference signal, the frames that are subjected to expansion and compression are identified. The binary codes are then extracted and hence the letters and the signature.

The scheme as proposed is a non-blind method as the original signal is required as a reference in order to determine the difference signal. Details of the watermark embedding process and the watermark recovery process are presented in Sections 4 and 5 respectively.

3. PSYCHOACOUSTIC MODEL

To ensure the imperceptibility of the embedded watermarks, it is necessary to understand the behavior of the Human Auditory System (HAS).

The hearing range for a normal individual is 20Hz to 20kHz. Typically, the HAS is most sensitive in the region from 1 to 9kHz, with the threshold in quiet the lowest at 3.3kHz. The threshold in quiet indicates the sound pressure level of a pure tone that is just audible. The audible threshold, however, changes in the presence of other audio signals. This is called masking effect. Masking is the phenomenon in which a sound normally audible in quiet becomes inaudible in the presence of another louder sound [1].

A psychoacoustic model is a mathematical model incorporating this masking phenomenon. The frequency spectrum of the hearing range is divided into a number of subbands. For Psychoacoustic Model I, the frequency spectrum of the hearing range is divided into 32 subbands. The model provides the masking threshold in each subband. Any sound below this threshold will be irrelevant to the human ears. Hence, if a spectral component with amplitude falling below this threshold is removed or added, there will be no perceptible difference to the listener. This is because these changes do not alter the spectral energy of the signal significantly [2].

For the watermarking scheme proposed, Psychoacoustic Model I for MPEG 2 coding [2,3] is adopted.

4. WATERMARK EMBEDDING PROCESS

The original audio signal is sampled at a sampling frequency (F_s) of 44.1 ksamples per second, and quantized with 16 bits per sample.

The digitized samples are segmented into frames of 1024 samples per frame. Successive frames have 512 samples overlapping. Each frame of audio samples is windowed with a Hanning window as shown in Fig.1.

The energy level of the signal is determined. The signature is embedded only if the frame has an energy level above a given threshold. For frames that are to be expanded or compressed, Fast Fourier Transform (FFT) is performed. The psychoacoustic model is then used to determine the masking threshold in each subband. The length of the frame of samples may be expanded by inserting samples with amplitude not exceeding the threshold of their respective subbands. Similarly, the length of the frame of samples may be shortened by

removing samples with amplitude not exceeding the threshold of their respective subbands.

Inverse Fast Fourier Transform (IFFT) is subsequently performed to recover the signal in the time domain. If additional samples are added in the frequency domain, there will be additional samples recovered in the time domain. Similarly, if samples are removed in the frequency domain, the number of samples in the time domain is decreased. In short, the number of samples added or removed in the frequency domain will translate into the extra or reduced number of discrete samples in the time domain.

The recovered frame of time domain samples is then added to the time domain samples of the frame preceding.

The choice of the number of samples to be slotted in or removed is a compromise between the quality of the watermarked signal and the recoverability of the watermark. After trial and error, it is assessed that the addition or removal of 4 samples per frame of 1024 samples is sufficient for coding purposes while at the same time the expanded or compressed frames do not add to perceptible distortion.

Fig 1. Concatenation of frames

In order not to alter the total length of the watermarked audio signal, the number of expanded frames is made equal to the number of compressed frames. A simple scheme is to have an expanded frame followed immediately by a compressed frame as indicated in Fig.1. After a set of compression and expansion, the altered signal is in synchronization with the original signal again. The pair of frames, which shall be referred to as Diamond frames, may be used to represent a binary 1 while a normal frame may be used to represent a binary 0.

5. WATERMARK RECOVERY PROCESS

The original digital audio samples are matched against the digital watermarked samples. The difference in amplitudes between the corresponding samples of the two sets is then obtained. A typical difference signal is shown in Fig.2.

The difference between the original samples and each pair of expanded and compressed (EC) frames shall normally take the shape of a diamond, while unmarked frames (normal frames) will return a difference that is

very close to zero. This is because there is a gradual increase in difference between the original samples and samples of the expanded frame with time and decrease in difference between the original samples and the samples of the compressed frame with time.

Fig 2 the difference signal

For the recovery process, the first task is to identify the Diamond frames as shown in Fig.2.

The diamonds are differentiated from the unmarked frames by the sloping sides. Theoretically, a diamond should have a slope greater than 0 while an unmarked frame should have a slope close to 0.

Once the binary symbols are identified, the next task is to determine the letters of alphabet and thence the signature from the binary sequence.

6. PERFORMANCE OF THE SYSTEM

The proposed system is tested using excerpts of the following five audio signals. 1) Classical Music: Summer in G minor by Vivaldi, 2) Instrumental Music: Sundial Dreams by Kevin Kern, 3) Heavy Instrumental Music: Livin' La Vida Loca by Ricky Martin, 4) Pop Music: It Must Have Been Love by Roxette, and 5) Male Speech.

Instrumental music, pop music, rock (heavy instruments) music, classical music, and speech (SS) were chosen for their varied characteristics that serve to test the versatility of the watermarking scheme.

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

The following five attacks were then conducted on the test watermarked signals.

1. Filtering

A Finite Impulse Response (FIR) low pass filter with cutoff frequency of passband at 10kHz, cutoff frequency of stopband at 11kHz, passband ripple 0.173dB and stopband attenuation 40dB is used.

2. Resampling

The test signal originally sampled at 44.1ksamples/s is re-sampled at 22.05ksamples/s, and then restored by sampling again at 44.1 ksamples/s.

3. Addition of White Gaussian Noise

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

4. Cropping

Some segments of the watermarked signal are replaced with segments of the signal attacked with filtering and additive noise.

5. MP3 Conversion

The watermarked test signal is encoded into a 128kbts/s MP3 format and then restored.

The watermark or signature embedded was 'FSW'. A signature is a complete string of three letters, for example 'FSW'. A 50% rate of signature recovery indicates that for a complete string of ten 'FSW's, five complete 'FSW's are recovered. The results of watermark recovery are presented in Tables 1 to 5.

Table 1 Rate of Recovery for Classical Music

Attack	Signature
Without attack	100%
Filtering	90%
Resampling	90%
Addition of noise	100%
Cropping	100%
MP3 Conversion	100%

Table 2 Rate of Recovery for Instrumental Music

Attack	Signature
Without attack	100%
Filtering	88.9%
Resampling	100%
Addition of noise	88.9%
Cropping	100%
MP3 Conversion	100%

Table 3 Rate of Recovery for Rock Music

Attack	Signature
Without attack	100%
Filtering	88.9%
Resampling	77.8%
Addition of noise	100%
Cropping	100%
MP3 Conversion	11.1%

Table 4 Rate of Recovery for Pop Music

Attack	Signature
Without attack	100%
Filtering	22.2%
Resampling	22.2%
Addition of noise	90%
Cropping	100%
MP3 Conversion	11.11%

Table 5 Rate of Recovery for Speech

Attack	Signature
Without attack	100%
Filtering	80%
Resampling	80%
Addition of noise	100%
Cropping	60%
MP3 Conversion	60%

Experimental results show that under attacks (put through additional signal processing), certain audio files like Instrumental Music and Classical Music stand out as good host signals for our watermarking scheme, in that the various attacks have not affected the watermark recovery significantly. Other files like Rock Music and Pop Music exhibit good percentage recovery of complete signatures under some attacks like the addition of noise and cropping, but are inadequate under others like filtering and MP3 compression. For Speech, cropping and MP3 Conversion affect the rate of recovery most.

In all cases under investigation, although the rate of recovery is affected after attacks, the owner can still be identified.

7. CONCLUSION

A novel watermarking method using time-frequency expansion and compression is proposed. Frames of samples are converted to frequency domain; samples are inserted or removed under the guide of Psychoacoustic Model I before conversion back to time domain. An expanded frame followed by a compressed frame is treated as one binary code while a normal frame is treated as the other binary code. The watermark in the form of binary code is embedded in sections of the signal with sufficiently high-energy contents.

For recovery, the difference between the watermarked signal and the time-aligned original signal is obtained. The difference is the greatest for the frames that are expanded or compressed. This fact is used to use to extract the binary code and hence the watermark.

Experimental results show that recovery is perfect for all signals that are not attacked. For signals that are attacked, the rate of recovery of the signature depends on the type of audio signal and the method of attack.

8. REFERENCES

- [1] Pohlmann, KC, Advanced Digital Audio, Carmel, 1991.
- [2] Watermarking research by DR Fabien A.P. Petitcolas <http://www.cl.cam.ac.uk/~fapp2/>
- [3] "Information Technology- Coding of Moving Pictures and Associated Audio for Digital Storage up to about 1.5Mbits/s", ISO/IEC IS 11172, 1993.
- [4] Bassia, P, Pitas, IP, "Robust Audio Watermarking In Time Domain", EUSIPCO 1998, 8-11 Sept., Patras, Greece, pp. 25-28.
- [5] Foo, SW, Yeo, TH, Huang, DY, An Adaptive Audio Watermarking System, In Proceedings of IEEE TENCON 2001, pp. 509-513.
- [6] Hecht, D, "Embedded Data Glyph Technology for Digital Hardcopy Documents", in Proc. SPIE, vol. 2171, 1995.
- [7] Kirovski, D, Malvar, H, "Embedding and Detecting Spread Spectrum Watermarks under the Estimation Attack", Microsoft Research.
- [8] Foote, J, Adcock, J "Time Base Modulation: A New Approach to Watermark Audio and Images", Palo Alto Laboratory, Inc.
- [9] Garcia, RA, "Digital Watermarking of Audio Signal using Psychoacoustic Auditory Model and Spread Spectrum Theory", University of Miami.
- [10] Broesch, JD, Digital Signal Processing Demystified, HighText Publications, 1997.
- [11] Smith, WW, Smith, JM, Handbook of Real-time Fast Fourier Transforms, IEEE Press, 1995.
- [12] Tempelaars, S, Signal Processing, Speech and Music, Swets & Zeitlinger Publishers, 1996.
- [13] Ingle, VK, Proakis, JG, Digital Signal Processing using Matlab, Brooks/Cole Publishing Company, 2000