A blind audio watermarking scheme using peak point extraction

Foo, Say Wei; Xue, Feng; Li, Mengyuan

2005

Foo, S. W., Xue, F., & Li, M. (2005). A blind audio watermarking scheme using peak point extraction. IEEE International Symposium on Circuits and Systems, ISCAS 2005, (pp. 4409-4412). Singapore: School of Electrical and Electronic Engineering.

https://hdl.handle.net/10356/90924

https://doi.org/10.1109/ISCAS.2005.1465609

A Blind Audio Watermarking Scheme Using Peak Point Extraction

Foo Say Wei, Xue Feng and Li Mengyuan
School of Electrical and Electronic Engineering
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798

Abstract -- **In this paper, a blind digital audio watermarking scheme is proposed. A psychoacoustic model is used to ensure the imperceptibility of the embedded watermark. The watermark is embedded in the Discrete Cosine Transform (DCT) domain by changing the phase of five information carriers determined using the psychoacoustic model. Blind recovery is achieved through the application of a novel synchronization scheme, the peak point extraction (PPE) scheme. Error correcting codes are used to enhance the robustness of the system. Informal listening reveals excellent imperceptibility of the embedded watermark. The watermarked signal was put through various forms of attacks. Experimental results show that recovery of watermark is perfect for signal not subjected to attack and the system is robust against attacks such as cropping and re-sampling.**

Index Terms -- **watermarking, frequency masking, digital signal processing, linear block codes, Discrete Cosine Transform**

I.INTRODUCTION

Digital watermarking techniques find useful applications in copyright protection and authentication. If the original signal is not required for the recovery of the watermark, it is called the blind watermarking scheme; and if the original signal is required, it is called the non-blind watermarking scheme.

Many schemes have been proposed [1-10]. For all the schemes, synchronization is one of the crucial elements in any blind watermarking scheme as the location of watermarked frames must be accurately determined during the recovery process. Most synchronization schemes fall into two categories. In one category, extra synchronization information is inserted into the audio signal to indicate the location of watermark. In the other category, special points with specific features are selected for watermark insertion. The former is suitable for all types of audio signals, but it may distort the original audio signals and draw the attention of attackers. The latter, in contrast, does not have such problems because no additional information is inserted into the audio signal. However, the features may vary among different types of audio signals.

II.PSYCHOACOUSTIC MODEL

In general, audio watermarking schemes rely on the imperfection of the human auditory perception [6] so that the embedded watermark is not audible. Psychoacoustics is the science that quantifies the human perception of sound. In the human hearing process, one phenomenon called *Frequency Masking* is the most important to ensure the imperceptibility of the embedded watermark. Frequency masking is the phenomenon where a sound (maskee) normally audible in quite becomes inaudible in the presence of a louder sound (masker).

A psychoacoustic model is a mathematical model that tries to imitate the human hearing mechanism. The psychoacoustic model incorporates the frequency masking phenomenon and estimates the masking threshold. Any sound below the masking threshold is inaudible to human ears.

Garcia [9] proposed an algorithm to estimate the masking threshold in the psychoacoustic model. This algorithm incorporates the critical band rate, the basilar membrane spreading function and the spectral flatness measure.

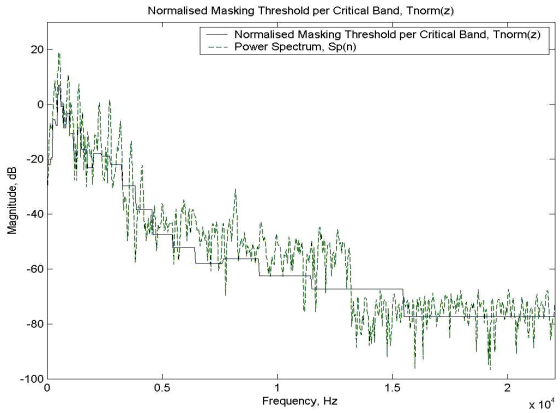

Fig.1 Normalized Masking Threshold

Fig.1 shows the normalized masking threshold estimated by the algorithm. The spectrum components that fall below the masking threshold are not readily perceived by the human auditory organs. Therefore, the audio signal will not be degraded if the watermark is embedded into the original signal using those components below the masking threshold. This provides the basic framework for watermark embedding.

III.THE PROPOSED SYSTEM

The flow-chart describing the essential steps of the proposed system is depicted in Fig. 2.

A. Peak Points Extraction

First, a set of synchronization points are identified in the time domain. A novel energy-feature-based synchronization scheme called the Peak Points Extraction (PPE) scheme is proposed in this paper. For this scheme, the power of the original signal is specially shaped by raising the sample value to a high power as exemplified by the following equation:

$$x'(n)=x^4(n) \quad (1)$$

where $x(n)$ is the original audio signal, and $x'(n)$ is the signal after the special shaping. Power of 4 is chosen as outstanding peaks can be easily identified.

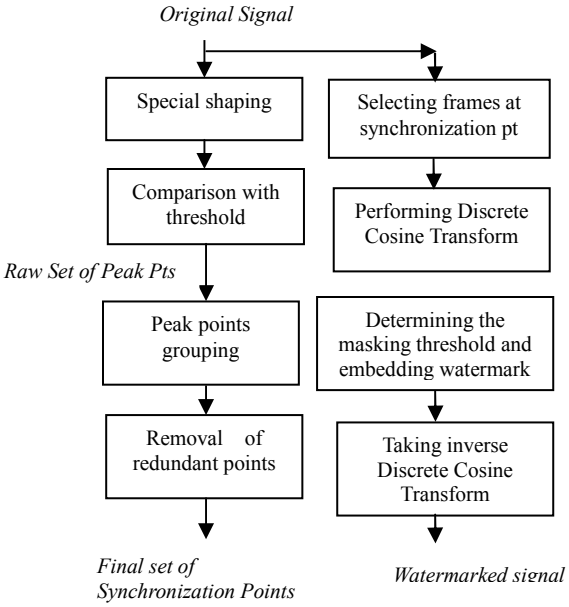

Fig.2 Block diagram of the proposed scheme

This process intentionally modifies the signal shape to exaggerate the energy differences between the peak regions and low-energy regions (Fig.3). After special signal shaping, the specific regions for watermark insertion are then identified by comparing the resulting signal with a threshold.

The threshold is set to be 15% of the sample value of the highest peak after special signal shaping. Samples that have value higher than the threshold are extracted as the peak points. A raw set of peak points is thus obtained.

The peak points usually appear in group consisting of many samples. Groups of peak points are first identified. The watermark is to be embedded between two consecutive groups of peak points. The last point of every group is taken as a synchronization point. As one frame is used to encode one bit of the watermark, we only consider consecutive groups of peak points when the number of samples between the groups is large enough to encode the watermark.

If the number of samples/frames between two consecutive peak points does not satisfy the requirement stated above, the last peak point is considered a redundant synchronization point, and is removed from the raw set of peak points. At the end of the process, a set of synchronization points is obtained.

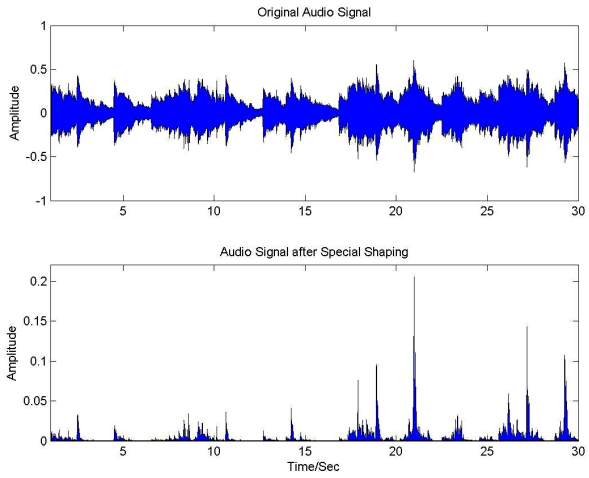

Fig.3 Audio signal before and after the special shaping

B. Alphanumeric Coding

The watermark, which represents the signature or the copyright information, is embedded into the audio signal in binary form. If the signature or the copyright information consists of alphanumeric characters, the characters are first encoded using binary codes before embedding. For the experiments reported in this paper, a 6-bit code is used to represent the alphanumeric characters (62 symbols: 10 decimal digits, 26 letters of alphabet in both upper case and lower case).

Error correcting codes [11] are well established for binary codes and they are employed for our watermark encoding. A (6,10) linear block code is adopted in this paper. Such a code is able to correct all single-bit errors, and up to 5 double-bit errors.

C. Watermark Embedding

The proposed audio watermarking scheme is carried out in the frequency domain. Embedding the watermark in the frequency domain has the advantages over embedding in the time domain in terms of imperceptibility and robustness [2]. Discrete Cosine Transform (DCT) is adopted as it is found to be suitable for audio signals [12].

The actual watermark embedding consists of three stages: pre-processing, layer one embedding and layer two embedding.

1) Pre-Processing: The alphanumeric symbols of the signature are encoded into 6-bit binary data words (without error correction) or 10-bit binary code words using (6,10) linear block codes.
2) Layer One Embedding(Symbol Embedding):The locations for embedding the alphanumeric symbols are identified.

Each synchronization point is the starting point of one embedded alphanumeric symbol.

The *n*-bit code word representing the alphanumeric symbol is embedded after the synchronization point using *n* consecutive non-overlap bit-embedding frames. A bit-embedding frame is referred to as a frame of samples used to embed one binary bit. The *n* consecutive bit-embedding frames are together referred as a symbol-embedding frame.

In most cases, the number of synchronization points is greater than the number of alphanumeric symbols in the signature. So the signature is repeatedly embedded throughout the audio signal. Such redundancy increases the accuracy of recovery.

3) Layer Two Embedding (Bit Embedding): After the symbol-embedding frames are determined, the locations of all bit-embedding frames are also confirmed. The binary bit '1' or '0' is embedded into the bit-embedding frames in the layer two embedding process.

Every bit-embedding frame has 1024 samples. The psychoacoustic model is applied to each bit-embedding frame. The mask threshold of each bit-embedding frame is estimated. The binary bit is embedded in the Discrete Cosine Transform (DCT) domain by manipulating those frequency components below the masking threshold, which are known as the imperceptible components. More robustness can be achieved if hidden data are placed among the low frequency coefficients in the DCT domain [10]. For each bit-embedding frame, the imperceptible component with the maximum power is selected from among all imperceptible components, and is used as a reference point. The reference point is always found in the low frequency range. Five successive imperceptible components after the reference point are used to code the binary bit. They are referred to as the information carriers. Five information carriers are used to add in redundancy to reduce the probability of error in the recovery process.

If a binary '1' is to be embedded in the bit-embedding frame, the signs of all five information carriers are made positive, regardless of the original signs. On the other hand, if a binary '0' is to be embedded in the bit-embedding frame, the signs of all five information carriers are made negative, regardless of the original signs.

This bit-embedding mechanism preserves the original power spectrum. The bit is embedded into the information carriers by changing their signs (the phases). The embedding process does not alter the amplitudes of those information carriers. The preservation of the power spectrum ensures that the reference point and the five information carriers can be determined without error or displacement in the watermark recovery process.

D. Watermark Recovery

The watermark recovery process is basically the reverse of the watermark embedding process. It also consists of three stages:

1) Pre-Processing: The primary task of the pre-processing stage is to recover the synchronization points. The synchronization points are extracted from the watermarked audio signal. Subsequently, the locations of all bit-embedding frames are determined. The recovery process starts with recovering the binary bit embedded in every bit-embedding frame.
2) Layer Two Recovery (Bit Recovery): A psychoacoustic model is applied to every bit-embedding frame. The masking threshold is estimated and the imperceptible components are determined. The reference point and information carriers are then determined in the same manner as in the embedding process.

The signs of the DCT coefficients of the five information carriers are examined. A positive sign indicates a binary '1' and a negative sign indicates a binary '0'. The decision is made based on the signs of the majority of the five information carriers.

3) Layer One Recovery (Symbol Recovery): Every symbol-embedding frame consists of *n* bit-embedding frames. After the bit-recovery process in the *n* consecutive bit-embedding frames, an *n*-bit binary code word is recovered from each letter-embedding frame. The alphanumeric symbols are then decoded and the watermark recovered.

IV. SYSTEM PERFORMANCE

To assess the performance of the proposed system, an imperceptibility test and a robustness test were performed. Some methods for advanced audio benchmarking can be found in [13].

Audio signals of 30 seconds each derived from five diverse sources are selected to test the performance of the proposed scheme. A (6,10) linear block codes with bit interleaving is used for error correction.

A. Imperceptibility Test

Informal listening using headset reveals that the watermark does not affect the quality of the original audio signals. This is as expected because the psychoacoustic model is incorporated into the scheme and the embedding process simply changes the phases of the spectral components. It is known that human auditory system is less sensitive to phase change. The imperceptibility of the embedded watermark is ensured.

B. Robustness Test

With no added signal processing, the recovery of the watermark is perfect for all five types of test signals. The effects of the following five types of attacks are then investigated.

1) Cropping: 10% of the audio samples are cropped and added back into the signal after adding colored noise.
2) Resampling: The audio signal is first down-sampled at 22.05 KHz, and then up-sampled at 44.1 KHz.

3) Addition of Noise: White Gaussian Noise is added to the audio signal to give 30 dB SNR.
4) Low-pass Filtering: The audio signal is filtered by a FIR lowpass filter with passband edge of 10 KHz.
5) MP3 Compression: The audio signals are compressed into MP3 format by MPEG-1 Layer-3 (LAME 3.92) encoder and decoded.

The signature (watermark) 'FSW' is employed as the testing signature. The experimental results of different attacks are shown in Table I. In the table, *ROCLR* stands for the Ratio of Correct Letters Recovered, while *ROCBR* stands for the Ratio of Correct Bits Recovered.

It can be seen that the system is able to withstand cropping and re-sampling very well. The ability to recover the signature is also dependent on the type of audio signal.

The ratios are computed based on the first signature embedded in the signal only. If repeated signatures in the signal are used and the majority rule applied, the accuracy may be further increased.

As the bit recovery rate is much higher, the performance of the system can further be improved by using error correcting codes with higher correction capability.

Table I
Performance of the System

Types of Attack	Types of Signal	ROCLR	ROCBR
Cropping	*Classic*	100%	100%
	Piano	100%	100%
	Instrumental	100%	100%
	Pop	88.2%	96.1%
	Speech	92.9%	98.8%
Re-sampling	*Classic*	83.3%	91.7%
	Piano	94.4%	99.1%
	Instrumental	75%	81.3%
	Pop	62.3%	84.1%
	Speech	80.7%	90.3%
Addition of noise	*Classic*	66.7%	91.7%
	Piano	22.2%	67.6%
	Instrumental	87.5%	97.9%
	Pop	58.8%	83.3%
	Speech	70.3%	82.7%
Low-pass filtering	*Classic*	33.3%	69.4%
	Piano	38.9%	78.7%
	Instrumental	50%	72.9%
	Pop	35.3%	70.6%
	Speech	75.0%	84.5%
MP3 compression	*Classic*	16.7%	41.7%
	Piano	55.5%	82.4%
	Instrumental	37.5%	70.8%
	Pop	47.1%	74.5%
	Speech	70.3%	81.7%

V. CONCLUSION

A blind digital audio watermarking scheme using Peak Position Extraction scheme is proposed. The scheme makes use of a novel approach, the Peak Point Extraction for synchronization. Watermark consisting of a short sequence of alphanumeric characters is repeatedly embedded in the Discrete Cosine Transform domain. The proposed system also makes use of a psychoacoustic model and linear block codes with bit interleaving to improve the performance of the system under various forms of attack.

Experimental results show that the proposed audio watermarking scheme possesses excellent imperceptibility. Listeners can hardly distinguish the original and watermarked audio signals. The proposed scheme also shows good robustness against attacks, especially cropping and re-sampling.

REFERENCES

[1] Swanson,M.D., Bin Zhu, Tewfik, "Current state of the art, challenges and future direction for audio watermarking", in Proceedings of IEEE International Conference on Multimedia Computing and Systems Vol.1, 1999, pp.19-24.

[2] C.P.Wu, P.C.Su, C.J.Kuo, "Robust Audio Watermarking for Copyright Protection", in Proceedings of SPIE's 44th Annual Meeting on Advanced Signal Processing Algorithms, Architectures, and Implementations IX (SD39), 1999.

[3] Bassia,P., Pitas,I., Nikolaidis,N.: "Robust audio watermarking in the time domain", in IEEE Transactions on Multimedia 3 (2001) pp. 232–241.

[4] W.Li, X.Y.Xue, and X.Q.Li, "Localized Robust Audio Watermarking in Regions of Interest", in Proceedings of ICICS-PCM 2003, 2003.

[5] Megias,D., Herrera-Joancomarti,J., Minguillon, J.: A robust audio watermarking scheme based on MPEG 1 layer 3 compression", in CMS 2003. LNCS 963, Springer-Verlag (2003) pp.226–238.

[6] H.J.Kim, Y.H.Choi, J.W.Seok, and J.W.Hong, "Audio Watermarking Techniques", Intelligent Watermarking Techniques, Chapter 8, pp.185-218, 2004.

[7] C. T. Hsieh and P. Y. Tsou, "Blind Cepstrum Domain Audio Watermarking Based on Time Energy Features", 14th Int. Conf. on Digital Signal Processing, pp705-708, Greece, 2002,

[8] J. W. Huang[1], Y. Wang[1], and Y. Q. Shi[2], "A Blind Audio Watermarking Algorithm with Self-Synchronization", ISCAS 2002, Scottsdale, Arizona, US, 2002.

[9] R.Garcia, "Digital Watermarking of Audio Signals Using a Psychoacoustic Auditory Model and Spread Spectrum Theory", in AES 107th Convention, New York, 1999.

[10] J.Cox, J.Kilian, T.Leighton, and T.Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", in IEEE Trans. on Image Processing, 6(12), 1997, pp.1673-1687.

[11] B.P.Lathi, "Modern Digital and Analog Communication System", Oxford University Press, 1998, pp.728-737.

[12] R.Aelinski, P.Noll, "Adaptive Transform Coding of Speech Signals", in IEEE Transaction on ASSP, Vol.25, 1979, pp.89-95.

[13]Dittman, J., Steinebach,M., Lang,A., Zmudzinski,S., "Advanced audio watermarking benchmarking", in Proceedings of the IS&T/SPIE's 16th Annual Symposium on Electronic Imaging, Vol.5306, Sant Jose, CA, USA.

[14] J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Trans. on Image Processing, 6(12), pp.1673-1687, 1997.

[15] L. R. Rabiner, and R. W. Schafer, "Digital Processing of Speech Signals", Prentice-Hall, Inc., pp.150-158, 1978.