World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:12, No:2, 2018

Robust and Transparent Spread Spectrum Audio Watermarking

Ali Akbar Attari, Ali Asghar Beheshti Shirazi

***Abstract*—**In this paper, we propose a blind and robust audio watermarking scheme based on spread spectrum in Discrete Wavelet Transform (DWT) domain. Watermarks are embedded in the low-frequency coefficients, which is less audible. The key idea is dividing the audio signal into small frames, and magnitude of the 6th level of DWT approximation coefficients is modifying based upon the Direct Sequence Spread Spectrum (DSSS) technique. Also, the psychoacoustic model for enhancing in imperceptibility, as well as Savitsky-Golay filter for increasing accuracy in extraction, is used. The experimental results illustrate high robustness against most common attacks, i.e. Gaussian noise addition, Low pass filter, Resampling, Requantizing, MP3 compression, without significant perceptual distortion (ODG is higher than -1). The proposed scheme has about 83 bps data payload.

***Keywords*—**Audio watermarking, spread spectrum, discrete wavelet transform, psychoacoustic, Savitsky-Golay filter.

I. Introduction

DIGITAL audio watermarking is one of the challengeable areas in the multimedia applications to protect multimedia contents against intelligent offending. Digital watermarking is a data-hiding technique embedding a sequence of information in host audio signal that can later be extractable while the quality of watermarked signal will be maintained. To preserve the quality of host signal, the signal manipulating must be perceptually inaudible. Also, the watermarked audio signal must be able to be robust against various kinds of attacks such as Gaussian noise addition, Low pass filter, Resampling, Requantizing, MP3 compression [6], [8].

Audio watermarking schemes encompass Spread Spectrum, Quantization Index Modulation, Echo hiding, Least Significant Bit Modification, Phase coding and Modulation, and Patchwork based [6], [14]. Audio watermarking is implemented in the time and frequency domain. The frequency domain methods have a higher robustness and impermeability, because of using the signal character and human perception specifications. On the other hand, the time domain schemes have a lower complexity, and they are easily implementable, but they have a low robustness [6]. Among these schemes, Spread Spectrum technique has several advantages such as high robustness, strong security, and easy achievable synchronization. But, it comes across with host interference which restricts the extraction performance [8]. Some researchers try to suppress this problem by improving the Spread Spectrum watermarking.

Malvar and Florêncio [11] proposed the improved SS (ISS) method. The ISS scheme could remove the partial host interference at the embedder and could improve the performance of the watermark extraction. Malik et al. [10] proposed a frequency-selective SS watermarking method which randomly selects subband signals of host audio for watermark embedding. The FSSS-based watermarking exploited the simultaneous masking property of the human the auditory system in watermark embedding. Li and Fang [7] proposed an adaptive normalization coefficients for correlation and adaptive decision thresholds for comparing in the detector. In this method, the detector can adjust to various audio signals and normal attacks. Kim et al. [15] proposed a selective correlation detecting method which uses a portion rather than the whole block of the signal for correlation. In this method, a better detection performance can be achieved by using a subset of blocks with higher watermark embedding power. Seok [13] proposed a time-domain Spread Spectrum approach based on independent component analysis (ICA) to estimate the embedded watermark, and then, the correlation-based detector is used. The estimation-correlation based watermark detector is used to improve the watermark-to-signal ratio (WSR) as a limitation of blind detection. In this particular method, the ICA as a blind source separation (BSS) method is employed to suppress the host signal interference at the detector. This method is not utilized perceptual characteristic of the watermarked audio. Li et al. [8] proposed a Spread Spectrum watermarking method and PCA extraction-based watermarking scheme before the correlation detector. This watermarking scheme exploits the perceptual analysis in watermark embedder and extractor to improve performance.

In order to make audio watermarking methods effective, basic requirements must be satisfied which are described as follows:

Imperceptibility: Inaudibility of difference between a watermarked audio signal and an original audio signal which preserves the quality of the audio signal after watermarking. Signal-to-Noise Ratio (SNR) and Objective Difference Grade (ODG) are used to evaluate the imperceptibility of a watermarked audio signal.

Robustness: After encountering various attacks, the watermarks must be extractable.

Capacity: The number of bits as a watermark can embed without degrading imperceptibility. This criterion is measured

A. A. Attari is with School of Electrical Engineering Iran University of Science & Technology (IUST), Narmak, Tehran 16844, Iran (phone: +98(71) 52338413; e-mail: A_Attari@vu.iust.ac.ir).

A. A. Beheshti Shirazi is with School of Electrical Engineering Iran University of Science & Technology (IUST), Narmak, Tehran 16844, Iran (corresponding author, phone: +98(21) 73225620; e-mail: abeheshti@iust.ac.ir).

in bits per second.

In audio watermarking, there is trade-off rule between basic requirements such as imperceptibility, robustness, and capacity [9].

If the algorithm is blind, the original signal may not need while extracting. Otherwise, the algorithm will be aware or non-blind [4].

The rest of paper is organized as follows. Section II introduces psychoacoustic model. Section III describes the proposed watermarking algorithm. In Section IV, the watermarking method is tested against most common attacks and evaluated regarding imperceptibility and capacity. Finally, conclusions are made in Section V.

II. Psychoacoustic Model

In spread spectrum watermarking technique, the watermark signal is added to the host signal as an additive noise. To keep the watermarks inaudible, we often use the minimum masking threshold (MMT) calculated from psychoacoustic model to shape the amplitude of watermark signal. ISO/IEC MPEG-1 Standard uses Psychoacoustic Model 1 to determine the MMT. Psychoacoustic Model 1 for Layer I is employed in our audio watermarking method because of its higher efficiency. In our method, the input to Psychoacoustic Model 1 is host frame, and the corresponding output is its MMT. The procedure of implementation is mentioned as follows [9]:

1) FFT analysis and SPL normalization
2) Identification of tonal and nontonal maskers
3) Decimation of invalid tonal and nontonal maskers
4) Calculation of individual masking thresholds
5) Calculation of global masking threshold
6) Determination of the MMT

III. Proposed Algorithm

In this paper, an audio watermarking method based upon the DSSS technique in the frequency domain is proposed. The scheme consists of two processes; embedding procedure and extraction procedure are illustrated in Fig. 1 and Fig. 2, respectively.

A. Embedding

Firstly, in the preprocessing block, framing and DWT is applied on the host audio signal to extract the 6^{th} level of approximation coefficients in each frame, then the 6^{th} level of approximation coefficients will be modified based on spread spectrum. Finally, the inverse DWT will be applied for regeneration of audio signal. Embedding steps are represented as follows:

1) Divide the host signal into frames. In our case, we use frame with one-second length.
2) DWT for giving the 6^{th} level of approximation coefficients will be applied to the host frame.
3) In watermark spreading block, the watermark $W_0 \in \{+1, -1\}$ is modulated by pseudo noise sequence(r_s) with length (l) to produce the modulated watermark W_m. To keep W_m inaudible, scaling factor α utilized to control the amplitude of W_m based on (1). Then, S_i' is obtained by adding W_m to S_i as host signal. Each watermark bit embeds in the frame repetitively with repetitive code value (d); in the other words, repetitive code denotes how much times a unique bit watermark embeds in host frame.

$$S_i' = S_i + \alpha W_m(i) \tag{1}$$

The scale factor (α) can control the strength of watermarking. Greater scale factor increases robustness against attacks and decreasing imperceptibility and vice versa. For a better perceptual quality, adaptive scale factor which is given in (2) is used.

$$\alpha = \beta \times P_s \tag{2}$$

where β is an adjustable parameter to change the balance between imperceptibility and robustness. In our case, we introduced 0.95 to keep energy distortion below the masking threshold according to method which is used in [5]. P_s is an index which is given based on (3), where T is the minimum masking threshold (MMT) for each frequency subframe which is calculated based on procedure of Section II.

$$\boldsymbol{P_s = \sqrt{T}} \tag{3}$$

4) Finally, inverse DWT was applied on S_i' for giving the watermarked signal.

B. Extraction

In the extraction process, watermark bits are extracted by using a linear correlation between the watermarked signal and pseudo noise sequence (r_s).

The extraction process can be summarized in the following steps:

1) By applying least squares Savitzky-Golay smoothing filters, the variance of the host signal is greatly decreased. And a reduced variance of host signal will result in a decreased extraction error rate [2].
2) Apply Framing and DWT to calculate the 6^{th} level of approximation coefficients of the watermarked audio.
3) The linear correlation between 6^{th} level of DWT approximation coefficients and r_s is calculated based on (4).

$$\boldsymbol{R_e(i) = \frac{1}{N}\sum_{j=1}^{N} S''_i(j).r_s(j) = \frac{1}{N}\sum_{j=1}^{N}[S_i(j) + \alpha W_m(i)].r_s(j)}$$
$$\boldsymbol{= \frac{1}{N}\sum_{j=1}^{N} S_i(j).r_s(j) + \frac{1}{N}\sum_{j=1}^{N} \alpha W_m(i) r_s(j)} \tag{4}$$

If the host frame S_i and the r_s are independent, the first term of (4) is close to zero. Otherwise, the second term has a large value and its sign might change the extracted watermark. Then, it is necessary that S_i and r_s are uncorrelated as much as possible.

The great role of applying Savitzky-Golay filter, which is employed to de-correlate the host audio signal and the watermark and degrade the interference from the host audio signal [12], as a preprocessing block is used to reduce the effect of the host audio signal to the fullest extent [9].

4) Each watermark bit W_e is extracted based on (5).

$$\boldsymbol{W_e(i)} = \begin{cases} \mathbf{1}, & \boldsymbol{if\ R_e(i) \geq 0} \\ \mathbf{0}, & \boldsymbol{if\ R_e(i) < 0} \end{cases} \tag{5}$$

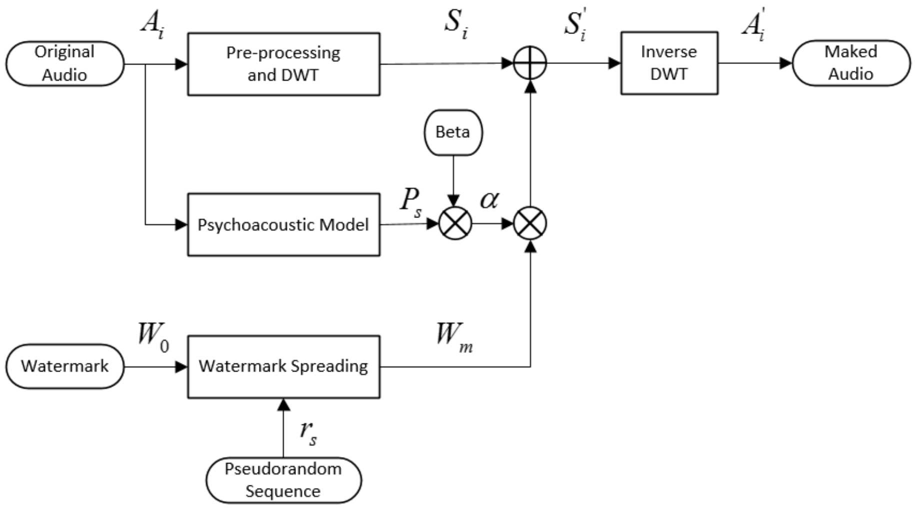

Fig. 1 Structure of Embedding Scheme

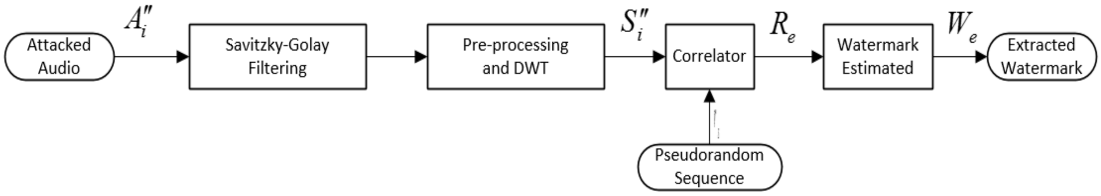

Fig. 2 Structure of extraction scheme

If the sign of R_e is positive, then the extracted watermark is one, and otherwise is zero. After extracted whole watermarks of each subframe, the output watermark can be given based on voting scheme. Based on this method, if the number of zero sample extracted in each subframe with length (d=6) is larger than three, the extracted watermark bit will be zero and if the number of one sample extracted in each subframe is larger than or equal to three, the extracted watermark bit will be one.

In the proposed method, the watermarks embed in low-frequency coefficients of DWT with controllable embedding strength by human ear's characteristic to achieve robustness and transparency. Using Savitzky-Golay filter will be improved the robustness of the proposed algorithm.

IV. Experimental Results

We use eight audio files of Pop music from album 'Rust by No, Really' to assess the performance and the effectiveness of the proposed algorithm. These audio files are sampled at 44.1 kHz with 16 bits quantization resolution and have two channels. The original audio files have an MP3 format which has convert to WAV format.

The audio signal is decomposed by using the Daubechies-8 wavelet (db8). Although the Daubechies filter presents nonlinear phase, it is the best choice for watermark embedding on the magnitude of coefficients. Then, it does not affect watermark embedding. Moreover, the Daubechies filter presents high-resolution and maximum smoothness capability in its pass-band. In the proposed method, there are adjustable parameters l, d for applying a balance between watermarking requirement: imperceptibility, robustness, and capacity. Increasing the pseudo noise sequence length (l) would cause lower capacity, however, it would also lead to higher correlation gain and robustness against attacks. On the other hand, increasing repetitive code value (d) cause to increase robustness as well as decrease capacity. We presume $l = 8, d = 6$ for the performance evaluation in this paper.

We use the ODG to measure the perceptual quality of the watermarked audio. ODG evaluation is known an effective tool for the comparative evaluation of audio quality. The results of ODG have range between -4 to 0. The watermarked audio file which has ODG value between -1 to 0, indicates that their audio quality is marked as imperceptible. Another objective test to measure modification error which is added to original audio signal after watermark embedding is SNR.

Table I demonstrates the measured ODG which is in the range of -0.58 to -0.96. All of the ODG values are higher than -1, indicating the embedding process achieve an excellent perceptual quality. SNR is about 20 dB which is indicated that watermark embedding is inaudible. The perceptual shaping process, as well as watermark embedding in 6^{th} level of DWT approximation coefficients (equal to frequency range 0 to 344 Hz) which have a low sensitivity of human's ear, guarantees the imperceptibility of the embedded watermark signal.

TABLE I
IMPERCEPTIBILITY TEST RESULTS

Audio Track	ODG	SNR
Beginning of the End	-0.58	18.03
Go	-0.73	18.16
Stop Payment	-0.87	19.93
Thousand Yard Stare	-0.85	19.12
Breathing on Another Planet	-0.69	19.18
Floodplain	-0.96	19.23
Citizen, Go Back to Sleep	-0.64	21.06
Molten	-0.74	20.37

For extraction performance test, the robustness of the proposed algorithm is evaluated by calculating BER of decoding under various attacks. BER is calculated based on the difference between a watermark extracted and watermark inserted. Attacks are applying based on the proposed algorithm applications. The most common attacks are described as follows:

1) Gaussian noise addition: adding additive white Gaussian noise with SNR = 30 dB.
2) Resampling: resampling the watermarked audio signal to 22020 Hz and then back to 44100 Hz.
3) Requantizing: requantizing the watermarked audio signal to 8 bits/sample and then back to 16 bits/sample.
4) Low-pass Filter: low-pass filtering with 4 kHz cut-off frequency.
5) MP3 compression: MP3 compressing the watermarked audio signal with a MPEG layer 3 coder implemented by LAME [3] at a 64 kbps bit rate and then decompressing.

As shown in Table II, the watermarked audio files under the different attacks have BERs equal to zero. Experimental results confirm strongly robustness against most common attacks such as Resampling, Requantizing, Gaussian noise addition attacks, MP3 compression, and Low-pass filter attacks. Since MP3 compression and Low-pass filter attacks change or eliminate high-frequency coefficients [1], these attacks could not affect watermark extraction beside the watermarks embedded in coefficients which have frequency lower than 344 Hz. In this way, the proposed algorithm is strongly robust against MP3 compression and Low-pass filter attacks. Also, the proposed algorithm presents robustness against Resampling, Requantizing, and Gaussian noise addition, because the watermark embedding in low-frequency subband of DWT has low temporal resolution compared to a high-frequency subband.

The proposed watermarking scheme obtained 83 bps data payload.

TABLE II
ROBUSTNESS TEST RESULTS

Audio Track	No attack	Resampling	Requantizing	Noise addition	Low-pass filter	MP3
Beginning of the End	0.00	0.00	0.00	0.00	0.00	0.00
Go	0.00	0.00	0.00	0.00	0.00	0.00
Stop Payment	0.00	0.00	0.00	0.00	0.00	0.00
Thousand Yard Stare	0.00	0.00	0.00	0.00	0.00	0.00
Breathing On Another Planet	0.00	0.00	0.00	0.00	0.00	0.00
Floodplain	0.00	0.00	0.00	0.00	0.00	0.00
Citizen, Go Back to Sleep	0.00	0.00	0.00	0.00	0.00	0.00
Molten	0.00	0.00	0.00	0.00	0.00	0.00

TABLE III
COMPARISON OF DIFFERENT SCHEMES IMPERCEPTIBILITY AND CAPACITY

Schemes	SNR	ODG	Payload
SS–ICA [13]	19.99	−0.76	43.07
SS [7]	23.84	-	43.07
SS [8]	28.00	-0.62	43.07
DWT-Fibonacci [1]	14.93	−0.62	686.00
Proposed	19.38	−0.76	83.00

Several exiting watermarking methods are compared to the proposed algorithm for a better evaluation. The imperceptibility and capacity results of different schemes are listed in Table III. The proposed method has a proper perceptual quality similar to others, while the capacity of proposed presented by the proposed method is two times higher than others.

TABLE IV
COMPARISON OF DIFFERENT SCHEMES ROBUSTNESS

Schemes	No attack	Resampling	Requantizing	Noise addition (dB)	Low-pass Filter (Hz)	MP3
SS–ICA [13]	0.07	2.70	0.90	2.86 (30 dB)	11.58 (4KHz)	2.22
SS [7]	0.27	17.87	3.36	0.69 (36 dB)	3.45 (11 KHz)	14.78
SS [8]	0.03	0.03	0.01	0.53 (20 dB)	0.03 (10 KHz)	0.29
DWT-Fibonacci [1]	0.00	0.00	0.00	0.00 (30 dB)	0.00 (4 KHz)	0.00
Proposed	0.00	0.00	0.00	0.00 (20 dB)	0.00 (4 KHz)	0.00

In Table IV, robustness results of various schemes are demonstrated. The SNR (dB) and cut of frequency (kHz) for different methods are shown in brackets. In SS-ICA algorithm, SNR is retained on 20 dB deliberately and the obtained acceptable ODG, since it uses LPC and noise shaping; however, one the major problems in SS-ICA method is not extracting watermarks correctly in no-attacks condition. The proposed algorithm provides a higher robustness against Gaussian noise addition, Resampling, Requantizing, Low-pass filter and Mp3 attacks compared to other algorithms. The proposed algorithm like DWT-Fibonacci [1] method embeds watermarks in 6^{th} approximation coefficients of DWT to achieve imperceptibility and robustness, simultaneously.

The results which are listed in Tables III and IV show that the proposed algorithm outperforms the others in robustness under partly similar imperceptibility and capacity.

V. Conclusion

In this paper, the robust and blind Spread Spectrum audio watermarking method based on the 6^{th} level of DWT approximation coefficients is proposed. Performance improvement is achieved by watermark embedding in low-frequency coefficients of DWT as well as exploiting the frequency-masking property of a human auditory system in watermark shaping. Also, using Savitzky-Golay filter before the correlation detector helps to decrease interference between host signal and watermark. The experimental results demonstrate that the proposed method has a high robustness against most common attacks such as Resampling, Requantizing, Gaussian noise addition, Low-pass filter, and MP3 comparison. The proposed algorithm provides a payload capacity of 83 bps. Imperceptibility result shows the proposed algorithm effectiveness in preserving the perceptual quality of watermarked audio. Future studies will be carried out the statistical analysis and the perceptual quality to achieve more robustness against other attacks.

References

[1] Attari, A. A. and Asghar Beheshti Shirazi, A. 2017. Robust and Blind Audio Watermarking in Wavelet Domain. Proceedings of the International Conference on Graphics and Signal Processing (New York, NY, USA, 2017), 69–73.
[2] Cvejic, N. 2004. Algorithms for audio watermarking and steganography. Oulun yliopisto.
[3] D. Ellis http://labrosa.ee.columbia.edu/matlab/.
[4] Fallahpour, M. et al. 2015. Audio watermarking based on Fibonacci numbers. Audio, Speech, and Language Processing, IEEE/ACM Transactions on. 23, 8 (2015), 1273–1282.
[5] He, X. 2012. Signal Processing, Perceptual Coding, and Watermarking of Digital Audio: Advanced Technologies and Models. Information Science Reference.
[6] Hu, H.-T. and Hsu, L.-Y. 2015. A DWT-Based Rational Dither Modulation Scheme for Effective Blind Audio Watermarking. Circuits, Systems, and Signal Processing. (2015), 1–20.
[7] Li, L. and Fang, X. 2010. Adaptive detection for spread spectrum audio watermarking. IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS). (2010), 58–62.
[8] Li, R. et al. 2016. Spread spectrum audio watermarking based on perceptual characteristic aware extraction. IET Signal Processing. 10, 3 (2016), 266–273.
[9] Lin, Y. and Abdulla, W.H. 2015. Audio Watermark - A comprehensive Foundation Using MATLAB. Springer.
[10] Malik, H. et al. 2008. Robust audio watermarking using frequency-selective spread spectrum. IET Information Security. 2, 4 (2008), 129–150.
[11] Malvar, H. S. and Florêncio, D. A. F. 2003. Improved spread spectrum: a new modulation technique for robust watermarking. IEEE transactions on signal processing. 51, 4 (2003), 898–905.
[12] Press, W. et al. 1987. Numerical Recipes: The Art of Scientific Computing.
[13] Seok, J. 2012. Audio watermarking using independent component analysis. Journal of information and communication convergence engineering. 10, 2 (2012), 175–180.
[14] Tewari, T. K. 2015. Novel Techniques for Improving the Performance of Digital Audio Watermarking for Copyright Protection. Ph.D. dissertation, Dept. Computer Science Engineering & Information Technology, Jaypee Institue Of Information Technology, Noida, India.
[15] Wook Kim, H. et al. 2010. Selective correlation detector for additive spread spectrum watermarking in transform domain. Signal Processing. 90, 8 (2010), 2605–2610.

Ali Akbar Attari received the B.Sc. degree in Broadcast Engineering from Islamic Republic of Iran Broadcasting University (IRIBU), Tehran, Iran in 2013 and M.Sc. degrees in Communication Engineering from Iran University of Science and Technology (IUST), Tehran, Iran in 2017. He has been collaborating with IRIB research and development department. His research interests include multimedia processing and Secure Communication.

Ali Asghar Beheshti Shirazi. received the B.Sc. and M.Sc. degrees in Communication Engineering from Iran University of Science and Technology (IUST) in 1984 and 1987, respectively and Ph.D. from Okayama University, Japan in 1995. In 1995, he joined the Department of Electrical Engineering, IUST, where he currently is an Assistant Professor. His research interests include Digital Image Processing and compression, Data Communication Networking and Secure Communication.