

Research Letter

Efficiently Synchronized Spread-Spectrum Audio Watermarking with Improved Psychoacoustic Model

Xing He¹ and Michael S. Scordilis²

¹BrainMedia LLC, New York, NY 10016, USA

²Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA

Correspondence should be addressed to Xing He, starchina@gmail.com

Received 15 November 2007; Accepted 1 February 2008

Recommended by Mark Liao

This paper presents an audio watermarking scheme which is based on an efficiently synchronized spread-spectrum technique and a new psychoacoustic model computed using the discrete wavelet packet transform. The psychoacoustic model takes advantage of the multiresolution analysis of a wavelet transform, which closely approximates the standard critical band partition. The goal of this model is to include an accurate time-frequency analysis and to calculate both the frequency and temporal masking thresholds directly in the wavelet domain. Experimental results show that this watermarking scheme can successfully embed watermarks into digital audio without introducing audible distortion. Several common watermark attacks were applied and the results indicate that the method is very robust to those attacks.

Copyright © 2008 X. He and M. S. Scordilis. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Watermarking is one of the most promising techniques to promote copyright protection and content authentication. Key to effective watermarking is the accurate and practical inclusion of human hearing and visual perception properties so that the embedded information remains imperceptible and the technique is robust to distortion and deliberate attacks.

Due to the high sensitivity of the human auditory system [1], audio watermarking is much more challenging compared to embedding watermarks into images. In the past decade, various audio watermarking techniques have been proposed including lowest-significant bit coding, phase coding, echo coding, and spread spectrum [2]. However, a suitable psychoacoustic model remains indispensable for any effective robust audio watermarking scheme.

Most psychoacoustic models employed for watermarking so far are related to the perceptual entropy (PE) [3, 4]. The short-time Fourier transform (STFT) is typically applied to provide a time-varying spectral representation of the signal [5–7]. Although it is adequate for stationary signals, it cannot provide detailed information for transient signals due to its fixed temporal and spectral resolution. Instead, audio sig-

nal characteristics would be analyzed and represented more accurately by a more versatile description, which would provide time-frequency multiresolution more suitable to that signal dynamics. The wavelet transform presents an attractive alternative by being able to provide resolution details that better match the hearing mechanism [8]. Specifically, long windows analyze low-frequency components and achieve high-frequency resolution while progressively shorter windows analyze higher-frequency components to achieve better time resolution. Such flexible and detailed signal representation, as provided by wavelet analysis, could contribute to effective watermarking as well, provided distortion remains inaudible while watermarking capacity remains considerable.

Implementing audio watermarking in the wavelet domain has only recently started been investigated. In [9], Wu et al. introduced an efficient self-synchronized audio watermarking scheme. However, no psychoacoustic model was included in their algorithms and, as a result watermarking transparency was only possible by using a user-adjustable watermark strength factor, which was experimentally set for different audio signals. Similar attempts [10, 11] also relied on a user-interactive approach for tuning the watermark. Such approach greatly limits the application of these techniques.

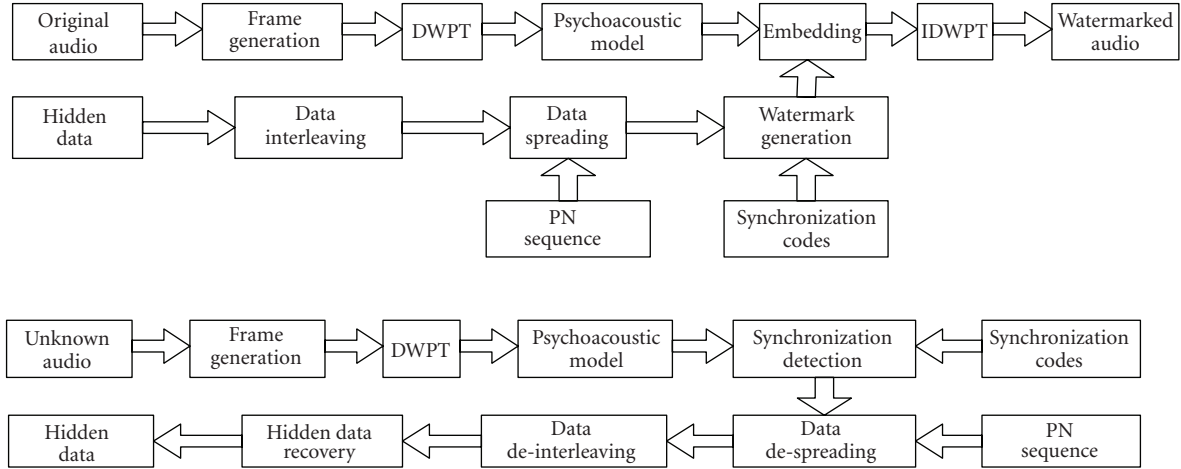


FIGURE 1: Proposed watermarking system: (a) encoder and (b) decoder.

Although wavelet analysis in psychoacoustic model computation has been recently explored [12–15], it is either computationally expensive [12, 13] by having to rely on the Fourier transforms for the computation of the psychoacoustic model itself, or their critical bands approximation deviate from the standard partition by varying degrees [14], which may result in objectionable audible distortion in the reconstructed signal.

In this paper, we propose an efficiently synchronized spread spectrum audio watermarking scheme based on a psychoacoustic model that uses the discrete wavelet packet transform (DWPT). This DWPT-based psychoacoustic model, first introduced in [15], boasts a wavelet packet-based decomposition that better approximates critical bands distribution, and it incorporates effective simultaneous and temporal masking, thus maintaining perceptual transparency and providing an attractive alternative to discrete Fourier transform (DFT)-based approaches for audio watermarking. An efficiently synchronized spread spectrum technique is used to embed watermarks by taking advantages of the proposed psychoacoustic model, and it achieves better watermarking robustness at higher payload capacity.

The paper is organized as follows. In Section 1, we briefly introduce the DWPT-based psychoacoustic model and its advantages over the DFT-based approaches from an audio watermarking perspective. The proposed watermarking system is described in Section 2 followed by the experimental procedures and results in Section 3. The conclusion is in Section 4.

1. DWPT-BASED PSYCHOACOUSTIC MODEL

While related techniques [12–14] share a similar general structure, the psychoacoustic model proposed in [15] achieved an improved decomposition of the signal into 25 critical bands using the discrete wavelet packet transform (DWPT). The result was a partition, which approximates the critical band distribution much closer than before as we

pointed out in [15]. Furthermore, the masking thresholds were computed entirely in the wavelet domain.

In [15], we evaluated and compared the proposed and the standard analysis methods from two useful perspectives: (1) the extent to which portions of the signal power spectrum can be rendered inaudibly and therefore provide space for audio watermarking without audibly perceived impact, and (2) the amount of reduction in the sum of signal-to-mask ratio (SSMR) that can be achieved, which is an indication of the degree with which the watermark robustness can be improved by embedding higher-energy watermark without introducing audible distortion.

The experimental results in [15] showed that the overall gain in the extent of masked regions provided by the proposed wavelet method is 20% and achieves SSMR reduction rate of 57%, indicating that a significant increase for watermark robustness is possible

2. PROPOSED WATERMARKING SYSTEM STRUCTURE

The proposed watermarking system consists of the encoder and decoder as illustrated in Figure 1.

The system encoder works as follows.

- The input original audio is segmented into overlapped frames, which are decomposed into 25 subbands by DWPT in wavelet domain as proposed in [16].
- Psychoacoustic model in [16] is applied to determine the masking thresholds for each subband.
- The data to be embedded (hidden data) are repeated and interleaved to enhance watermarking robustness.
- The interleaved data are spreaded by a pseudorandom sequence (PN sequence).
- Synchronization codes are attached at the beginning of the spreaded data, thus producing the final watermarks to be embedded.
- The watermarks are embedded into the original audio by meeting the masking thresholds constraints.
- Inverse DWPT (IDWTP) is applied on the above data to get the watermarked audio.

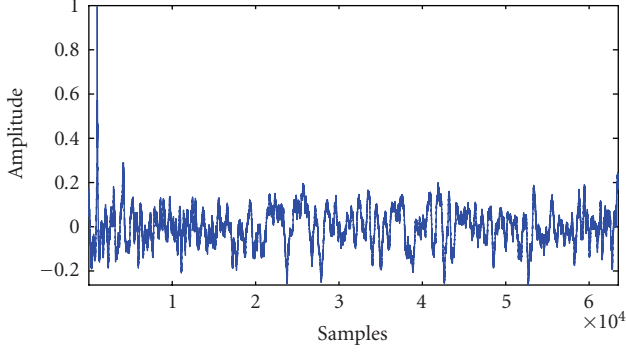


FIGURE 2: Synchronization code detection.

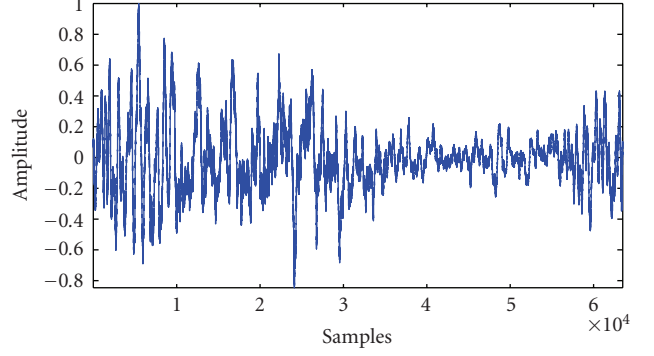


FIGURE 3: Damaged synchronization code.

The system decoder works in a reverse manner as the encoder, and it is also illustrated in Figure 1.

2.1. The synchronization process

Similar to most of the spread-spectrum-based watermarking systems, which require good synchronization between encoder and decoder, the proposed system also required synchronization in order to recover the watermarks. Synchronization is achieved by attaching synchronization codes before each spreaded watermark.

The synchronization code we employed here is a PN sequence and it is used to locate the beginning of the hidden data. Such enhance the ability to fight against desynchronization attacks like random cropping or shifting. During the decoding process, a fast search in limited space is performed by a matched filter to detect the existence of the synchronization code.

The fast search of synchronization code is possible due to the properties of the DWPT outlined as follows [9].

Suppose that $\{a_j\}$ is the original audio, M is one frame within $\{a_j\}$, N is another frame 2^k samples shifted from M , with the same length, $c_j^{k,M}$ and $c_j^{k,N}$ are the j th wavelet coefficients of M and N after k -level DWPT, then

$$c_{j+1}^{k,M} = c_j^{k,N}, \quad (1)$$

except for less than $L + 2$ boundary coefficients, where L is the length of the wavelet filter [9]. Therefore, in order to find one synchronization code within the watermarked audio, the decoder needs to perform at most 2^k times ($k = 8$ in our case) sample by sample searches instead of D times sample by sample searches, where D is the length of the spreaded watermark ($D > 60,000$ samples in our case). This results in greatly enhanced efficiency.

A typical result of the search is shown in Figure 2 where the peak denotes the start position of the synchronization code, in which case perfect synchronization between the watermark encoder and the decoder is achieved. However, after some attacks, the watermarks may get partially damaged resulting in obscured localization peaks, as shown in Figure 3 where the damage is severe.

Since our goal is to recover the watermark with as few errors as possible and since multiple watermarks may be em-

bedded in the audio file, it is better to skip seriously damaged watermarks and recover the watermark from the less damaged ones. In order to decide how bad the watermark has been damaged after some attack, we define a factor called “pratio” as

$$\text{pratio} = \frac{\max(o)}{\sum |o|/N}, \quad (2)$$

where o is the output of the detection filter, $|o|$ is the magnitude of o , and N is the length of o . Only when $\text{pratio} >$ threshold, the watermarked audio frame is considered not seriously damaged, and watermark recovery is performed on this frame.

2.2. Watermark embedding and extracting

In order to survive signal processing and malicious attacks, especially compression that damages most of high-frequency information, we embed the watermarks in the perceptually significant components of the audio signal, which lie predominantly in low-frequency areas.

The embedding process involves calculating the masking threshold for low-frequency bands and spreading the watermark with PN sequence.

In order to achieve higher robustness against additive noise, the original watermark information is repeated several times [16] and interleaved by an array similar to [17] and later spreaded by using the PN sequence.

Suppose the spreaded data are $\{m_i\}$, consisting of “1” second and “-1” second, for each frame, then the embedding rule is

$$c_k = \begin{cases} c_k + \alpha * m_i & \text{if } c_k^2 > T, \\ \sqrt{T} * m_i & \text{if } c_k^2 \leq T, \end{cases} \quad (3)$$

where c_k is the absolute value of the k th wavelet coefficient of low-frequency components in the frame, α is a factor ($0 \leq \alpha \leq 1$) used to control the watermark strength, m_i is the symbol to be embedded in this frame, and T is the masking threshold for that low-frequency subband. Increasing α typically improves watermarking system robustness by embedding higher-energy watermark, at the risk of producing perceptual distortion.

TABLE 1: Proposed watermarking system bit error rate.

Attack type	Country (%)	Jazz (%)	Pop (%)	Rock (%)	Classic (%)	AVG (%)
(a)–(d)	0	0	0	0	0	0
MP3/64 Kbps	0.78	0.71	0.52	0.27	0.19	0.49
MP3/32 Kbps	2.03	1.61	3.33	0.27	1.44	1.74
MP3/20 Kbps	2.52	7.37	5.94	2.08	13.44	6.27

In the decoding phase, the input signal is first segmented into overlapping frames, and the masking thresholds for the subband in each frame are calculated. Let $d = \sum c_k$, where c_k is the k th wavelet coefficient of low-frequency components in the frame, and it satisfies $c_k^2 \leq T$, where T is the masking threshold for that low-frequency subband. Then, the recovery decision rule is

$$w = \begin{cases} 1 & \text{if } d > 0, \\ -1 & \text{if } d \leq 0. \end{cases} \quad (4)$$

Data w are then despread by the same PN sequence and then deinterleaved by the same array. Since the watermarks are repeatedly embedded in the audio file, the majority rule in (5) is used to construct the final watermark from the individually recovered watermark. Suppose that N watermarks are individually recovered and the length of each watermark is D , then the k th symbol of the final recovered watermark is

$$W_k = \text{sign} \left(\sum_{i=1}^N w_{i,k} \right), \quad (5)$$

where $w_{i,k}$ is the k th symbol of the i th recovered watermark ($1 \leq k \leq D$), and sign is the function defined as

$$\text{sign}(k) = \begin{cases} 1 & \text{if } k > 0, \\ -1 & \text{if } k \leq 0. \end{cases} \quad (6)$$

By using the majority rule, we can usually recover the final watermark perfectly even if some individual watermarks are not error free.

3. EXPERIMENTAL PROCEDURES AND RESULTS

Experiments were conducted to evaluate the proposed psychoacoustic model followed by the test of the complete watermarking system.

A set of five audio files was used to evaluate the robustness of the proposed watermarking scheme. They contained varying musical pieces of CD-quality that included jazz, classical, pop, country, and rock music. The total of 160 bits information is embedded for a watermark bit rate of 8 bps. Several attacks were applied individually on the proposed watermarking system and they included the following.

- (a) *Random cropping*: samples are randomly deleted or added to the watermarked audio.
- (b) *Noise addition*: white noise with -36 dB power level compared to the watermarked audio is added.

(c) *Resampling*: the watermarked audio is down sampled to 22.05 kHz and then up sampled to 44.1 kHz.

(d) *DA/AD conversion*: the watermarked audio is played on the computer, and the output is recorded through the line-in jack on the sound card of the same computer.

(e) *MP3 compression*: the watermarked audio is compressed into MP3 format at various bit rates and then decompressed back into wave file.

From the results shown in Table 1, we can see that the proposed watermark scheme is quite robust to such attacks. Watermarks are recovered perfectly after the attacks like random cropping, noise addition, resampling, and DA/AD conversion. It also shows good robustness to MP3 compression, even at extremely low -20 Kbps bit rate.

Although recovery errors occurred after MP3 compression, as shown in Table 1, multiple watermarks are inserted into the audio file, and the final recovered watermark is based on a majority rule of the individual recovered watermarks. In all cases, the final recovered watermark contained no error even if several individual watermarks contained errors.

Subjective listening tests were conducted as well, and they confirmed that when embedding the watermark with the proposed method the processed audio signals were indistinguishable to the original for the proposed technique, resulting in a transparent watermark scheme.

4. CONCLUSION

In this paper, we have presented an audio watermarking method with an improved spread-spectrum technique and an enhanced psychoacoustic model that is based on the DWPT. The proposed method includes superior synchronization, which makes it possible to resynchronize after several attacks. The psychoacoustic model used in this watermarking system calculates the masking and auditory thresholds more accurately than other techniques, and it completes that in the wavelet domain and it renders watermarking transparent. It also provides broader masking capabilities compared to DFT-based psychoacoustic models thus revealing that larger signal regions are in fact inaudible and therefore providing more space for watermark embedding without noticeable effects. Furthermore, the signal-to-masking ratio is additionally reduced thus permitting stronger watermarks to be embedded, which, when combined with optimal signal location-picking, increases watermark robustness considerably.

REFERENCES

- [1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Academic Press, San Diego, Calif, USA, 2002.
- [2] W. Bender, D. Gruhl, N. Morimoto, and A. Liu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3-4, pp. 313–336, 1996.
- [3] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, vol. 66, no. 3, pp. 337–355, 1998.
- [4] S. Jung, J. Seok, and J. Hong, "An improved detection technique for spread spectrum audio watermarking with a spectral envelope filter," *ETRI Journal*, vol. 25, no. 1, pp. 52–54, 2003.
- [5] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.
- [6] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, 1988.
- [7] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
- [8] R. Polikar, "The wavelet tutorial," <http://users.rowan.edu/polikar/Wavelets/wtpart1.html>, May 2007.
- [9] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 69–76, 2005.
- [10] N. Cvejic and T. Seppanen, "Robust audio watermarking in wavelet domain using frequency hopping and patchwork method," in *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (ISPA '03)*, vol. 1, pp. 251–255, Rome, Italy, September 2003.
- [11] H. O. Kim, B. K. Lee, and N. Lee, "Wavelet-based audio watermarking techniques: robustness and fast synchronization," <http://citeseer.ist.psu.edu/644572.html>, September 2007.
- [12] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3463–3479, 1993.
- [13] N. R. Reyes, M. R. Zurera, F. L. Ferreras, and P. J. Amores, "Adaptive wavelet-packet analysis for audio coding purposes," *Signal Processing*, vol. 83, no. 5, pp. 919–929, 2003.
- [14] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1622–1635, 1999.
- [15] X. He and M. S. Scordilis, "An enhanced psychoacoustic model based on the discrete wavelet packet transform," *Journal of the Franklin Institute*, vol. 343, no. 7, pp. 738–755, 2006.
- [16] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 3, pp. 1345–1348, Salt Lake City, Utah, USA, May 2001.
- [17] R. A. Garcia, "Digital watermarking of audio signals using a psychoacoustic model and spread spectrum theory," in *Proceeding of the 107th Audio Engineering Society Convention*, New York, NY, USA, September 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

