AN INFORMED UNIPOLAR SPREAD SPECTRUM MODULATION FOR SELF-SYNCHRONIZED ROBUST WATERMARKING

Peng Zhang, Shuzheng Xu, and Huazhong Yang

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
E-mail: zhangp00@mails.tsinghua.edu.cn, {xusz, yanghz}@tsinghua.edu.cn

ABSTRACT

This paper describes an informed unipolar spread spectrum (SS) modulation to improve the performance of SS-based watermark extraction with self-synchronization. Compared with bipolar SS, unipolar SS, realized by employing a pair of orthogonal codes to represent the watermark bits, is more robust against the distorted correlation between the received watermarked signal and the reference. The informed embedding utilizes the host interference to improve the performance of unipolar watermark synchronization and decoding. Experiments on audio signals show that the proposed scheme is robust against most of the attacks while preserving the media fidelity, and effective self-synchronization is achieved.

Index Terms— Robust watermarking, spread spectrum, improved spread spectrum, informed embedding, unipolar modulation

1. INTRODUCTION

Digital watermarking is a technique that embeds a signature or additional data into the host media. It has become a solution to copyright protection, distribution control, broadcast monitoring, and other aspects of multimedia security. Among the existing schemes, spread spectrum (SS) modulation is an important category for robust watermarking [1]. It has the properties of interference immunity, low embedding distortion, and invariance to amplitude scaling. Moreover, SS-based watermarking has the ability of self-synchronization, in which no extra signal is needed to synchronize the embedded watermark with the reference. However, traditional SS watermarking suffers from inherent host interference caused by additive embedding and blind extraction, resulting in loss of the watermark robustness [2].

Many schemes have been proposed to improve the robustness of SS-based watermarking by reducing the host interference at the receiver or at the transmitter [3, 4]. Utilizing the host information at the transmitter is also known as informed embedding, which was inspired by modeling watermarking as communications with side information [2]. Improved SS (ISS) is one of the informed strategies for SS-based watermarking [3]. It utilizes the host projections on the watermark at the embedder to compensate for the host interference at the decoder. Besides the host rejection or cancellation methods, schemes such as the sign embedder [5] and the double-sided embedding [6] were proposed to utilize the host interference for watermark detection. However, their methods can not distinguish the information bits and are not applicable for watermark extraction or decoding.

This work focuses on SS-based watermark extraction with self-synchronization. As stated below, bipolar SS watermark extraction is vulnerable to false peaks of the distorted correlations between the received watermark and the reference. Therefore, a unipolar SS modulation is proposed and accompanied with an informed embedding strategy to improve the performance of watermark synchronization and decoding.

2. SS AND ISS WATERMARKING

Let $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{n}$ be the column vector expressions of the host signal, a pseudo-noise (PN) sequence with values of $+\sigma_u$ or $-\sigma_u$, and the channel noise, respectively. Vector length N determines the watermark robustness and the payload. The watermark bit to be embedded is mapped to one of the bipolar values of $+1$ and -1, and then denoted by b. The watermarked signal and the corrupted one under additive attack noise are represented by $\mathbf{s}$ and $\mathbf{y} = \mathbf{s} + \mathbf{n}$, respectively.

In traditional SS watermarking, information bits are embedded by $\mathbf{s} = \mathbf{x} + b\mathbf{u}$, with an embedding distortion of σ_u^2. For correlation-based blind extraction, the watermark decoder calculates the decision variable as

$$r = \langle \mathbf{y}, \mathbf{u} \rangle / \langle \mathbf{u}, \mathbf{u} \rangle = b + x + n \tag{1}$$

where $\langle \mathbf{y}, \mathbf{u} \rangle \triangleq \mathbf{y}^T \mathbf{u} / N$ is the inner product of vectors $\mathbf{y}$ and $\mathbf{u}$; $x \triangleq \langle \mathbf{x}, \mathbf{u} \rangle / \langle \mathbf{u}, \mathbf{u} \rangle$ and $n \triangleq \langle \mathbf{n}, \mathbf{u} \rangle / \langle \mathbf{u}, \mathbf{u} \rangle$ are the projections of vectors $\mathbf{x}$ and $\mathbf{n}$ on vector $\mathbf{u}$, respectively. The watermark bit is then decoded by using the decision rule $\hat{b} = \mathrm{sign}(r)$. Assume that the elements in vectors $\mathbf{x}$ and $\mathbf{n}$ are identical and independently Gaussian distributed, and obey $N(0, \sigma_x^2)$ and $N(0, \sigma_n^2)$, respectively. The signal-to-noise ratio (SNR) is measured by σ_x^2/σ_n^2. Then projections x and n are Gaussian random variables [3]:

$$x \sim N(0, \sigma_x^2/(N\sigma_u^2)), n \sim N(0, \sigma_n^2/(N\sigma_u^2)). \tag{2}$$

Note from (1) that the host signal acts as strong interference in the decision variable and thus deteriorates the decoding performance. As an informed embedding strategy, ISS reduces the host interference at the transmitter [3]. Its linear form can be written as $\mathbf{s} = \mathbf{x} + (\alpha b - \lambda x)\mathbf{u}$, where parameters α and λ control the embedding distortion and the removal of the host interference. To obtain the same distortion as in traditional SS, α is derived as $\alpha = \sqrt{1 - \lambda^2\sigma_x^2/(N\sigma_u^2)}$. At the watermark decoder, the decision variable is calculated as

$$r = \langle \mathbf{y}, \mathbf{u}\rangle / \langle \mathbf{u}, \mathbf{u}\rangle = \alpha b + (1-\lambda)x + n, \tag{3}$$

and the decision rule is the same as in traditional SS. For $\lambda \approx 1$, ISS reduces the host interference and thus outperforms traditional SS in decoded bit error rate (BER).

3. DRAWBACKS OF BIPOLAR SS WATERMARKING

Both traditional SS and ISS belong to bipolar SS modulation, in which the information is carried by the sign of the decision variable. Prior to the decoding process, the embedded PN sequence should be aligned with the reference to produce a correct decision output. Typical synchronization methods for bipolar SS watermarking detect the peak amplitude (maximum absolute value) of the correlation between the received signal and the reference, and then the embedded bit is recovered from the sign of the peak. Therefore, a sharp correlation between the received PN sequence and the reference is required. However, under some manipulations, especially over a frequency-selective channel, such a sharp correlation is deteriorated. Correlations between the received PN sequences and the reference are illustrated in Fig. 1, where several commonly used audio processing operations are applied. It can be seen that the channel manipulations reduce the peak amplitude of the correlation and increase the amplitude of the sidelobes[1]. Under additive noises or other attacks, the amplitude of one of the sidelobes may exceed that of the true correlation peak, resulting in misalignment between the received sequence and the reference. We call such a sidelobe a false correlation peak. It is important to note that a false correlation peak can also indicate the existence of the watermark but will mislead a bipolar watermark decoder to produce an error if the false peak has a sign opposite to that of the true one. This fact is well illustrated in Fig. 1(d) where the correlation contains several large sidelobes with a wrong polar. Moreover, under some circumstances such as sound transmission via the sound card we tested in Fig. 1(c), all samples of the watermarked signal are inverted, resulting in imperceptible changes but all inverted decoding results.

One way to cope with the above problems is to use unipolar SS modulation, in which the sign of the decision variable does not carry any information. Thus the decoder will not be affected by the false correlation peak with a polar opposite to that of the true peak.

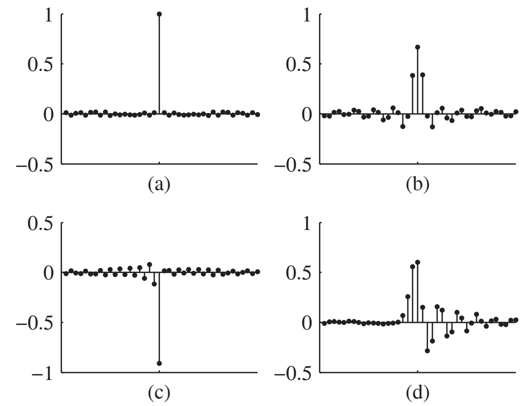

Fig. 1. Normalized correlation between the processed PN sequence and the reference ($N = 8192$). (a) no processing; (b) MP3 compression at $96\,\mathrm{kbps}$; (c) playing and recording via sound card; (d) lowpass filtering with a cutoff frequency of $12\,\mathrm{kHz}$. Only a few samples around the peak are plotted.

4. PROPOSED SCHEME

Unlike traditional SS and ISS modulation, in which a unique PN sequence is modulated by the watermark bits mapped to bipolar values of ± 1, our method employs two orthogonal sequences, $\mathbf{u}_0$ and $\mathbf{u}_1$, each with N elements of $+\sigma_u$ or $-\sigma_u$, to represent the unipolar watermark bits 0 and 1, respectively. The orthogonality ensures zero interference between the two sequences, i.e.,

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} \sigma_u^2, & i = j, \\ 0, & i \neq j, \end{cases} \; (i, j = 0, 1). \tag{4}$$

Let $b \in \{0, 1\}$ be the watermark bit to be embedded, and $\bar{b} = 1 - b$ be the alternative value of b. The informed unipolar embedding process can be formulated as

$$\mathbf{s} = \mathbf{x} + \alpha \mathrm{sign}(x_b)\mathbf{u}_b - x_{\bar{b}}\mathbf{u}_{\bar{b}}, \tag{5}$$

where $x_i \triangleq \langle \mathbf{x}, \mathbf{u}_i \rangle / \langle \mathbf{u}_i, \mathbf{u}_i \rangle$ $(i = 0, 1)$ is the projection of host vector $\mathbf{x}$ on watermark vector $\mathbf{u}_i$. Under Gaussian assumptions in (2), the average embedding distortion is $\alpha^2\sigma_u^2 + \sigma_x^2/N$, which is the same as that of ISS with $\lambda = 1$, and α can be derived as $\alpha = \sqrt{1 - \sigma_x^2/(N\sigma_u^2)}$ to obtain the same distortion level σ_u^2 as in traditional SS and ISS. Then the signal-to-watermark ratio (SWR) can be measured by σ_x^2/σ_u^2.

At the watermark extractor, two decision variables are calculated on the received signal:

$$r(i) = \left| \frac{\langle \mathbf{s} + \mathbf{n}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \right|, \; i = 0, 1. \tag{6}$$

[1]For a PN sequence with a good auto-correlation property, the cross-correlations in Fig. 1 reflect the channel impulse responses.

Substituting the embedding equation (5) and the orthogonality (4) into (6), we obtain

$$r(i) = \begin{cases} |\alpha + |x_b| + n_b|\,, & i = b, \\ |n_{\bar{b}}|\,, & i = \bar{b}, \end{cases} \tag{7}$$

where $n_i \triangleq \mathrm{sign}(x_i)\langle \mathbf{n}, \mathbf{u}_i\rangle \,/\, \langle \mathbf{u}_i, \mathbf{u}_i\rangle$ obeys the distribution of noise projection in (2). Then the watermark bit is recovered by using the decision rule

$$\hat{b} = \arg\max_{i=0,1}\{r(i)\}. \tag{8}$$

Note that the unipolar watermark decoder does not rely on the sign of the correlation peak, and allows either positive or negative sidelobes with large amplitude. Moreover, equation (7) exhibits some interesting behaviors of the host interference after the informed embedding. The host interference, x_b, on decision variable $r(b)$ increases the peak amplitude of the correlation between the received watermarked signal and the reference sequence $\mathbf{u}_b$, implying an improvement over ISS in watermark synchronization. On the other hand, the host interference on decision variable $r(\bar{b})$ is completely removed by using ISS at the embedder, and the gap between $r(b)$ and $r(\bar{b})$ is enlarged by $|x_b|$, resulting in better decoding performance than if x_b is cancelled.

5. PERFORMANCE ANALYSIS

5.1. Self-synchronization

Typical self-synchronization schemes for SS-based watermarking utilize the cross-correlation between the received signal and the reference PN sequence. When the amplitude of the correlation exceeds a predefined threshold, the embedded watermark is considered to be detected and aligned with the reference. A false positive error happens when the amplitude of the correlation exceeds the threshold but the received signal contains no watermark or the watermark has not been aligned with the reference. Under Gaussian assumptions, the detection probability P_D and the false positive probability (for each unsynchronized code shift) P_{FP} of the proposed scheme can be approximated to

$$P_D \approx 2\int_0^\infty f_x(x)\mathrm{Q}\left((\gamma - \alpha - x)\sqrt{N\sigma_u^2/\sigma_n^2}\right)dx \tag{9}$$

$$P_{FP} \approx 4\mathrm{Q}\left(\gamma\sqrt{N\sigma_u^2/(\sigma_x^2 + \sigma_u^2 + \sigma_n^2)}\right) \tag{10}$$

where γ is the predefined threshold, $f_x(x)$ is the probability density function of the host projection in (2), and $\mathrm{Q}(x) \triangleq 1/\sqrt{2\pi}\int_x^\infty \exp(-t^2/2)dt$.

Figure 2 illustrates the detection probability as a function of the false positive probability, often known as the receiver operating characteristic (ROC), for traditional SS, ISS, and the proposed scheme. Due to the absolute value of the host projection in (7), the proposed scheme exhibits higher correlation peaks than ISS, and better performance of synchronization or detection is achieved.

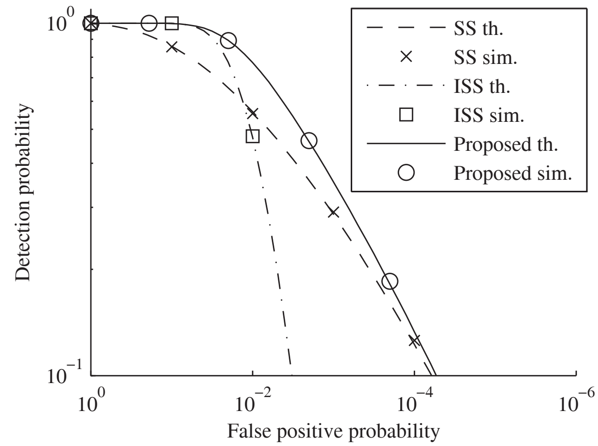

Fig. 2. Theoretical (th.) and simulated (sim.) ROC curves for SS, ISS ($\lambda = 1$), and the proposed scheme. $N = 8192$, $\mathrm{SWR} = 30\,\mathrm{dB}$, and $\mathrm{SNR} = 10\,\mathrm{dB}$.

5.2. Decoding robustness against additive noise

The robustness of watermark decoding against additive white Gaussian noise (AWGN) was evaluated. Figure 3 illustrates the BER performance for several schemes under different values of SWR and SNR. The results show that a great improvement over traditional SS is achieved, and over noise-free channel the embedded information can be completely recovered without errors. As a unipolar modulation, the proposed scheme performs a little worse than ISS under low SWRs. However, due to utilization of the host interference, the performance gap between them is reduced as SWR increases. Therefore, the proposed scheme is potentially appropriate for the applications where high media fidelity is required.

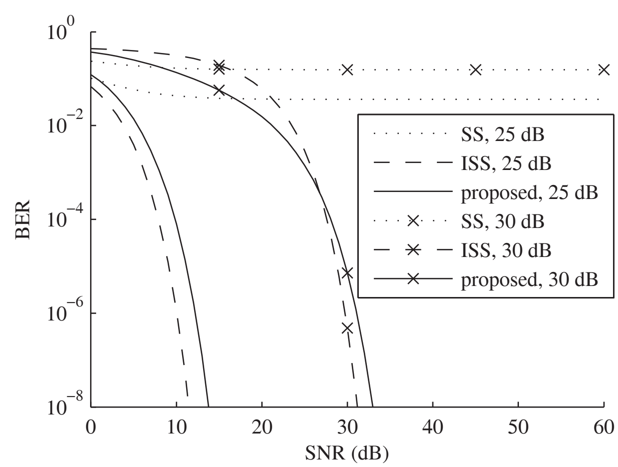

Fig. 3. BER vs. SNR under different SWRs (25 and 30 dB), for SS, ISS ($\lambda = 1$), and the proposed scheme. $N = 1024$.

6. EXPERIMENTS ON REAL AUDIO SIGNALS

The presented scheme was realized and evaluated on real audio signals. The hosts were from one hundred $44.1\,\mathrm{kHz}$ sampled, 16-bit quantized stereo audio sequences, each with a duration of 10 seconds. The watermark payload was $5.4\,\mathrm{bps}$ per channel for $N = 8192$. At the embedder, two bipolar walsh codes are selected as the orthogonal sequences, and then multiplied by another PN sequence to improve their auto-correlation and cross-correlation characteristics while preserving the orthogonality. Then the two sequences are shaped by using the psychoacoustic model in [7] to obtain inaudible watermarks. To avoid large compensation signals and suppress the peak power of the embedding distortion, the host audio is filtered by using a highpass filter (HPF) with a cutoff frequency of $3\,\mathrm{kHz}$ when calculating the host projections. At the decoder, the watermarked audio is filtered by using the same HPF and then correlated with the two reference sequences to recover the embedded watermark bits.

A search-lock strategy commonly used in SS systems was employed to obtain more stable synchronization. The search process finds the correlation peaks, and checks whether they are valid and reliable over several frames to avoid occasional false positive errors. When synchronization has been successfully established, the lock stage begins. In this stage, an accidental failure in detecting the correlation peak does not affect the synchronization state, unless it happens consecutively for several frames. Under a proper threshold, no false positive errors occurred during searching the unwatermarked audio; when synchronizing the watermarked audio, only one loss happened due to a long silence in the test audio clips. Thus, effective and robust self-synchronization is achieved.

To evaluate the watermark transparency, two objective measures, the SWR and the objective difference grade (ODG) ranging from -4 to 0, were calculated. Generally, a larger SWR or ODG implies better audio quality. After watermark embedding, the average SWR and ODG were $29.1\,\mathrm{dB}$ and -0.54, respectively. Double-blind subjective listening tests with five participants were also performed on ten of the audio clips, and no audible distortions were found. Therefore, high audio fidelity is obtained.

In robustness tests, BERs under the following attacks were measured: 1) no attack; 2) AWGN with an SNR of $20\,\mathrm{dB}$; 3) amplitude scaling (AS) to 200% and 10%; 4) lowpass filtering (LPF) with a cutoff frequency of $8\,\mathrm{kHz}$; 5) MP3 compression at the bit rate of $48\,\mathrm{kbps}$ per channel; 6) resampling (Resamp.) down to $22.05\,\mathrm{kHz}$, and then back to $44.1\,\mathrm{kHz}$; 7) requantizing (Requant.) to $8\,\mathrm{bits}$; 8) adding echoes with a delay of $100\,\mathrm{ms}$ and a decay of 40%. The results are shown in Table 1. It can be seen that the proposed method is robust against most of the attacks. The performance under AWGN attack is not as good as expected, because some low volume segments experience much lower SNRs than the specified average value and thus bring in most of the errors.

Table 1. Bit error rate (%) under typical attacks

No attack	0	AWGN $20\,\mathrm{dB}$	5.27
AS 200% / 10%	0/0	LPF $8\,\mathrm{kHz}$	0.12
MP3 $48\,\mathrm{kbps/ch}$	0.39	Resamp. $22.05\,\mathrm{kHz}$	0.05
Requant. $8\,\mathrm{bit}$	0.41	Echo $100\,\mathrm{ms}$, 40%	0.03

7. CONCLUSIONS

An informed unipolar SS modulation was presented for SS-based watermark extraction with self-synchronization. The performance is improved in two ways. First, compared with bipolar SS watermark extraction, unipolar SS is more tolerant of large sidelobes of the correlation between the received watermarked signal and the reference. Second, the informed embedding utilizes the host signal to increase the peak amplitude of the correlation and thus improves the performance of unipolar watermark synchronization and extraction.

8. REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shanmoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[2] I. J. Cox, M. L. Miller, and A. L. Mckellips, “Watermarking as communications with side information,” *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, July 1999.

[3] H. S. Malvar and D. A. F. Florêncio, “Improved spread spectrum: A new modulation technique for robust watermarking,” *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[4] N. Gerek and M. K. Mihcak, “Generalized improved spread spectrum watermarking robust against translation attacks,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, Jan. 2008, pp. 1673–1676.

[5] N. Merhav and E. Sabbag, “Optimal watermark embedding and detection strategies under limited detection resources,” *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 255–274, Jan. 2008.

[6] J. Zhong and S. Huang, “Double-sided watermark embedding and detection,” *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 297–310, Sept. 2007.

[7] ISO/IEC 11172-3, “Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s-part 3: audio,” Oct. 1993.