A New Informed Spread Spectrum Embedding for Robust Audio Watermarking

Peng Zhang, Ye Li, Jingsai Jiang, Yanhong Fan, Qiuyun Hao, Xiaofeng Ma
Shandong Provincial Key Laboratory of Computer Networks,
Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan, China
zhangp@sdas.org

***Abstract*—This paper presents a new embedding strategy to extend the performance bound of spread spectrum (SS) based watermarking by introducing more imperceptible distortions measured in the mean square errors (MSE). The potential of the host is sufficiently exploited and utilized to maximize the watermark robustness. This strategy is then realized in audio watermarking by adaptively inverting the host according to the correlation between the host and the modulated watermark sequence. The MSE embedding distortions can reach up to twice of the host power, while the perceptual distortions are effectively controlled. By utilizing more available host, the proposed method exhibits much better robustness than existing SS-based watermarking.**

Keywords—watermarking; data hiding; informed embedding; spread spectrum

I. INTRODUCTION

Rapid development of digitalized storage and transmission technologies makes it easier to distribute and counterfeit the multimedia works, resulting in challenges to multimedia security. Digital watermarking is a technique that can embed extra information into multimedia contents, usually in an imperceptible but robust way [1]. The embedded data can then be recovered for the purposes of copyright protection, integrity verification, broadcast monitoring, covert communication, etc. Generally, the watermark can be either detect-only [2, 3] or extractable [4]. This work focuses on the latter which is also referred to as data hiding.

Due to the capability of reliable transmission with low signal level, spread spectrum (SS) modulation is inherently a promising approach to robust and transparent watermarking [1]. However, the host media becomes strong interference to the watermark if simple additive embedding is applied [5, 6]. This problem is always an active topic of SS-based watermarking. According to the embedding strategies regarding the relation between the host and the watermark, we divide the SS-based watermark embedding into three generations.

1) Host-ignored SS. This category is also known as the blind (or non-informed) SS embedding in which the host is ignored and acts as noise. Traditional SS-based methods [1] belong to this type.

2) Host-rejected SS. Although the host signal usually dominates the overall interference to watermarking systems, it is known to the embedder, and thus can be cancelled at the embedding stage [5, 6]. This strategy is well known as the informed embedding [7, 8]. For SS watermarking, Malvar proposed an informed method named improved SS (ISS), in which the host interference is reduced by introducing a linear compensation signal to the embedder [9-11].

3) Host-utilized SS. In recent years, it has been found that the host does not always interfere with the watermark; in some cases it will contribute to watermark recovery and could be utilized rather than rejected [2-4]. To realize this concept in extractable watermarking, the correlation-aware ISS (CAISS) [4] and its special form called semi-ISS [12] were proposed. Their idea is to determine at the embedder whether the host signal tends to produce a decoding error, by checking the correlation value and the bit to be embedded. Then the host that may cause errors is reduced by using ISS, while leaving the innocuous host by using traditional SS. With the help of part of the host, better decoding robustness is achieved [13-15].

From the above roadmap, at least two trends for SS-based watermark embedding can be found: i) the correlation between the host and the modulated watermark should be as large as possible under the distortion constraint; ii) the host could be more sufficiently exploited to improve this correlation. For these purposes, in this paper we propose a new informed SS embedding strategy, and then realize it in audio watermarking.

II. EXISTING INFORMED SS EMBEDDING

By summarizing the existing methods such as traditional SS, ISS and CAISS, we present a generalized model of informed embedding for additive SS-based watermarking. The watermarked signal can be written as

$$\mathbf{s}=\begin{cases}\mathbf{x}+\alpha_1 b\mathbf{u}-\lambda_1\langle\mathbf{x},\mathbf{u}\rangle\mathbf{u}, & b\langle\mathbf{x},\mathbf{u}\rangle\geq 0\\ \mathbf{x}+\alpha_2 b\mathbf{u}-\lambda_2\langle\mathbf{x},\mathbf{u}\rangle\mathbf{u}, & b\langle\mathbf{x},\mathbf{u}\rangle<0\end{cases} \tag{1}$$

where $\mathbf{x}$ is the host vector, $b=\pm 1$ is the watermark bit to be embedded, and $\mathbf{u}$ is a pseudo-noise (PN) sequence with N elements of ± 1; α_1 and α_2 scale the watermark power; λ_1 and λ_2 control the strength of the compensation signal; and $\langle\mathbf{x},\mathbf{u}\rangle\triangleq\mathbf{x}^{\mathrm{T}}\mathbf{u}/N$ represents the normalized inner product of column vectors $\mathbf{x}$ and $\mathbf{u}$. When extracting the embedded data

from the received signal that is corrupted by noise **n**, the decoding rule is

$$\hat{b} = \text{sign}(r) \tag{2}$$

where $r \triangleq \langle \mathbf{s}+\mathbf{n}, \mathbf{u} \rangle$ is the decision variable at the decoder. The above notations will be used throughout the paper.

By adjusting the compensation signal, several typical SS embedding methods can be derived, as listed in Table 1. Note that although the compensation signal can reduce the host interference on the decoder, it introduces additional embedding distortions. Therefore, compared with ISS that aims to compensate for the overall host, more sophisticated methods like CAISS do not reject the host unless necessary. However, about half of the host still needs to be reduced. In this paper, we attempt to exploit the potential of this part.

TABLE I. TYPICAL SS SCHEMES DERIVED FROM (1)

Parameter Settings	Scheme	Feature
$\lambda_1 = \lambda_2 = 0, \alpha_1 = \alpha_2$	traditional SS [1]	host-ignored
$\lambda_1 = \lambda_2 > 0, \alpha_1 = \alpha_2$	ISS [9]	host-rejected
$\lambda_1 = 0, \lambda_2 > 0$	CAISS [4]	host-utilized

III. A NEW INFORMED SS EMBEDDING STRATEGY

Before describing our method, we first revisit the dirty paper coding (DPC) theory [6]. It declares that, for a watermarking system modeled as

$$\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{x} + \mathbf{w} + \mathbf{n}, \quad \| \mathbf{w} \|^2 \triangleq \langle \mathbf{w}, \mathbf{w} \rangle \le \sigma_w^2 \tag{3}$$

where **w** and **y** represent the watermark and the received signal, respectively, its ideal performance is determined only by σ_w^2 / σ_n^2, i.e. the power ratio of the embedding distortion to the channel noise, regardless of the power of the host. For SS-based watermarking, this performance bound has been approached by using ISS or CAISS. Then what can we do for a better embedding? Note that the embedding distortion in (3) is measured in terms of the mean square error (MSE). This is reasonable for traditional communications, but not always for multimedia watermarking that requires in fact the constraint of *perceptual* distortions. Based on this fact, we propose a new embedding strategy to circumvent the performance limitation. The embedding process is modeled as:

$$\mathbf{s} = H(\mathbf{x}) + \alpha b \mathbf{u} \tag{4}$$

where α is a scale factor; and $H(\mathbf{x})$ is a processing method performed on the host, which should meet three conditions:

a) $\langle H(\mathbf{x}), b\mathbf{u} \rangle \ge 0$;

b) the perceptual difference between $H(\mathbf{x})$ and **x** is limited or controllable;

c) the MSE distortion on the host, i.e., $\| H(\mathbf{x}) - \mathbf{x} \|^2$, is considerable compared with the watermark power.

Condition *a)* ensures that $H(\mathbf{x})$ does not interfere with the watermark under the decoding rule in (2); and the larger the correlation is, the better robustness can be obtained. Condition *b)* is obviously required for transparent embedding. Compared with the model in (3), the equivalent watermark in (4) can be written as $\mathbf{w} \triangleq \mathbf{s} - \mathbf{x} = H(\mathbf{x}) - \mathbf{x} + \alpha b \mathbf{u}$. Then condition *c)* will result in a much higher MSE value of $\| \mathbf{w} \|^2$. According to the DPC theory, the optimal performance that is related to $\| \mathbf{w} \|^2$ could be improved.

The idea behind the proposed strategy is to improve the performance bound by introducing more *imperceptible MSE distortions* to the host. Note that this idea is conceptual, and only provides a possibility to exploit the host for a better embedding. Currently, we have not found a common method to design such $H(\cdot)$; instead, we only attempt to realize it in audio watermarking.

IV. REALIZATION IN AUDIO WATERMARKING

Human auditory system (HAS) is insensitive to the absolute phases of audio signals. For example, inverting each sample of an audio segment will not bring in noticeable distortions. Based on this fact, we propose an informed SS embedding method called watermark-oriented host inversion (WOHI-SS) to realize the above strategy in audio watermarking.

A. Ideal Case

In WOHI-SS method, the host vector is inverted according to the correlation between the host and the modulated watermark sequence. Ideally, the host processing method in (4) is designed as

$$H(\mathbf{x}) = \text{sign}(bx) \cdot \mathbf{x} \tag{5}$$

where $x \triangleq \langle \mathbf{x}, \mathbf{u} \rangle$, and the embedding process can be written as

$$\mathbf{s} = H(\mathbf{x}) + \alpha b \mathbf{u} = \text{sign}(bx) \cdot \mathbf{x} + \alpha b \mathbf{u}. \tag{6}$$

It is easy to verify that $\langle H(\mathbf{x}), b\mathbf{u} \rangle = | x | \ge 0$; and the expected MSE distortion is $E(\| H(\mathbf{x}) - \mathbf{x} \|^2) = 2 \| \mathbf{x} \|^2$ (about half of the host vectors are inverted), which is twice as large as the host power. Therefore, conditions *a)* and *c)* for $H(\mathbf{x})$ are easily met.

Attention should be paid to condition *b)*. Although no audible difference can be found within the inverted audio segments, annoying distortions will arise at the boundaries between adjacent inverted and non-inverted frames. This is due to abrupt phase changes from 0 to π or vice versa that destroy the phase continuity. Fortunately, these distortions are only concentrated around the frame boundaries, and are relatively limited in the whole audio duration if long frames are used.

However, they still need to be reduced to improve the perceptual quality for practical use, as described below.

B. Practical Implementation

Abrupt changes in the relative phases between two adjacent audio frames will cause noises that sound like periodic beats. To solve this problem, we present a method to smooth the phase distortions while inverting most of the audio samples. The steps for $H(\mathbf{x})$ are described as follows.

1) Perform the ideal inversion of the host audio frames according to (5), and the smoothing operations below are performed only on the frame whose relative phase has been changed to $\pm\pi$ (i.e., the current frame has been inverted whereas the previous one has not, or vice versa).

2) Calculate the spectrum of current frame by applying the L-point short-time Fourier transform (STFT). The STFT blocks are hanning-windowed and 3/4 overlapped. Thus the N-point audio frame is divided into $4N/L$ blocks. Then phase transitions between 0 and π are realized by slowly changing the phase spectrum of each block. Denote the phase increment of the ith block by $\Delta\varphi_i$ that is defined as

$$\Delta\varphi_i = \begin{cases} \pi\sin(\dfrac{i}{2K}\pi), & i = 0,\cdots,K-1 \\ \pi, & i = K,\cdots,4N/L-1. \end{cases} \tag{7}$$

It is then added to the phase spectrum of the ith block.

3) Reconstruct the time domain frames by performing the inverse STFT with the new phase spectrum of each block.

The above methods of windowing and phase smoothing are illustrated in Fig. 1. Then the host inversion in (5) can be approximated with less noticeable phase distortions. Note that the samples in transition blocks cannot be completely inverted, and thus the resulting $H(\mathbf{x})$ for some frames may not meet the condition *a)* in section 3. In this case, ISS is employed to cancel the residual host interference. Then an approximated WOHI-SS embedding method is formulated as

$$\mathbf{s} = \begin{cases} H(\mathbf{x}) + \alpha b\mathbf{u}, & \langle H(\mathbf{x}), b\mathbf{u}\rangle \geq 0 \\ \mathbf{x} + \beta b\mathbf{u} - x\mathbf{u}, & \langle H(\mathbf{x}), b\mathbf{u}\rangle < 0 \end{cases} \tag{8}$$

where β controls the strength of ISS embedding.

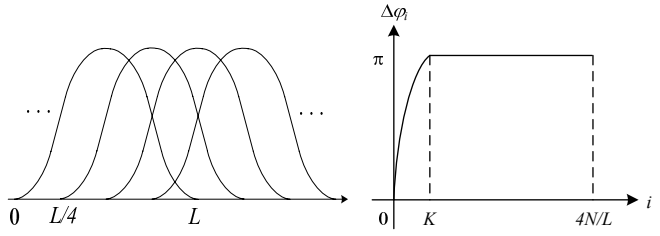

Fig. 1. Illustration of windowing (left) and phase smoothing (right).

V. Performance Evaluations

A. Theoretical Analysis

The theoretical performance is evaluated under Gaussian assumptions, i.e., the samples in host vector $\mathbf{x}$ and noise vector $\mathbf{n}$ are assumed to be Gaussian distributed. Then their projections on the watermark sequence can be written as $x \triangleq \langle \mathbf{x}, \mathbf{u}\rangle \sim \mathcal{N}(0, \sigma_x^2/N)$ and $n \triangleq \langle \mathbf{n}, \mathbf{u}\rangle \sim \mathcal{N}(0, \sigma_n^2/N)$ [9], where $\mathcal{N}(\mu,\sigma^2)$ denotes the Gaussian distribution with mean value μ and variance σ^2; σ_x^2 and σ_n^2 are the variances of the host and the noise, respectively.

For the ideal WOHI-SS method in (6), the decision variable at the decoder can be written as

$$\begin{aligned} r &\triangleq \langle \mathbf{s}+\mathbf{n}, \mathbf{u}\rangle = \mathrm{sign}(bx)\cdot x + \alpha b + n \\ &= (\alpha + \mathrm{sign}(bx)\cdot bx)b + n \\ &= (\alpha + |x|)b + n. \end{aligned} \tag{9}$$

Note that the magnitude of r can be increased by the host, resulting in better robustness than if the host does not exist or has been cancelled. Under the above assumptions, the decoding bit error rate (BER) can be derived as

$$P_b = 2\int_{-\infty}^{-\sqrt{N\alpha^2/\sigma_n^2}} Q\left(\frac{\sigma_n}{\sigma_x}z + \sqrt{\frac{N\alpha^2}{\sigma_x^2}}\right) f_0(z)dz - Q\left(\sqrt{\frac{N\alpha^2}{\sigma_n^2}}\right) \tag{10}$$

where $f_0(\cdot)$ is the probability density function of standard normal distribution, and $Q(\cdot)$ denotes the Q-function. The derivation is omitted here to save space.

Similarly, for the approximated WOHI-SS in (8), the decision variable can be written as

$$r \triangleq \langle \mathbf{s}+\mathbf{n}, \mathbf{u}\rangle = \begin{cases} \big(\langle H(\mathbf{x}), b\mathbf{u}\rangle + \alpha\big)b + n, & \langle H(\mathbf{x}), b\mathbf{u}\rangle \geq 0 \\ \beta b + n, & \langle H(\mathbf{x}), b\mathbf{u}\rangle < 0 \end{cases} \tag{11}$$

where $H(\mathbf{x})$ has been implemented in section IV.B. As in the ideal form, the host in the approximated method can also help improve the watermark robustness, and no decoding error occurs over noise-free channel.

Monte Carlo simulations are conducted to verify the performance of the proposed method. The existing methods including traditional SS, ISS and CAISS are compared under the same watermark payload and power. By adjusting the scale factors, the host-to-watermark power ratio (HWR) is set to be 35 dB, and the frame length is set to be $N = 8192$. For approximated WOHI-SS, we set $L = 64$ and $K = 32$ for an acceptable perceptual distortion and sufficient host inversion. The decoding performance is measured in BER as a function of E_b/N_0, where $E_b/N_0 \triangleq \mathrm{WNR}\cdot N/2 = N\alpha^2/(2\sigma_n^2)$ represents the watermark-to-noise power ratio (WNR) per bit.

The simulation results are shown in Fig. 2, where the markers correspond to the simulated values and the curves are obtained analytically (except the approximated WOHI-SS). The ideal DPC is equivalent to SS embedding into the host of zero power, whose analytic BER is $Q(\sqrt{2E_b/N_0})$. It can be seen that the ideal WOHI-SS method exhibits much better robustness than the others. Its approximated realization performs worse than the ideal form, but still better than ISS and CAISS in most cases.

Note that, although the proposed method may perform even better than the ideal DPC, this improvement does *not* mean a breakthrough to the DPC theory. In fact, the MSE distortions in the proposed method will be far stronger than in the others, even though they may sound acceptable. According to the DPC theory, increasing the MSE distortion while preserving the perceptual host quality will provide an opportunity towards the unexpected optimal performance.

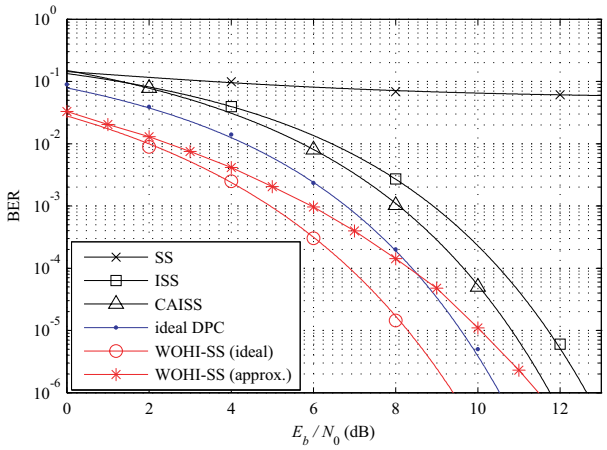

Fig. 2. Decoding performance comparisons under Gaussian assumptions.

B. Experiments

1) Perceptual Quality

In prior SS-based watermarking, the embedding distortions are only caused by watermark insertion, and usually appear as additive noises. Differently, the proposed strategy may introduce extra distortions caused by the host processing $H(\mathbf{x})$. To evaluate such distortions, we apply the WOHI-SS method to 100 audio segments of different types and styles. The 5-level objective difference grade (ODG) is selected to measure the audio quality (0: excellent; –1: good; –2: fair; –3: poor; –4: bad). The distributions of ODG for both ideal and approximated WOHI-SS are illustrated in Fig. 3, where the average values are –1.82 and –0.52, respectively. Due to the phase smoothing operation in approximated WOHI-SS, the perceptual distortions caused by host inversion can be effectively controlled.

The perceptual quality of approximated WOHI-SS is also evaluated through listening tests. For most audio segments, the embedding distortions are nearly inaudible. However, for some audio types such as piano solos, perceptible defects can be found. This may be because the audio with plenty of harmonics is more sensitive to the phase dispersion caused by the changes of relative phases. By adjusting parameters L and K, the perceptual quality and the decoding robustness can be balanced.

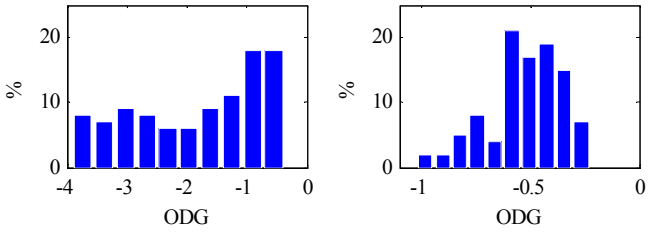

Fig. 3. ODG distributions of the two WOHI-SS methods.

2) Robustness

For robustness evaluations, the following commonly-used attacks are applied to the watermarked audio:

- AWGN: additive white Gaussian noise with relative power of –30 dB;
- LPF: lowpass filtering with cut-off frequency of 12 kHz;
- MP3: MP3 compression at the bit rate of 128 kbps;
- AAC: Advanced audio coding at the bit rate of 128 kbps;
- Requantization: requantization from the resolution of 16 bit to 8 bit;
- Resampling: resampling from 44.1 kHz to 22.05 kHz, and then back to 44.1 kHz.

Several existing SS-based embedding methods are also tested and compared with the proposed approximated WOHI-SS. To obtain similar perceptual audio quality, the average ODG for each method is controlled to be around –0.5, by adjusting the sequence power. By setting $N = 8192$, the watermark payload is set to be 5.4 bps for sampling frequency of 44.1 kHz. The results are listed in Table 2, indicating that the proposed method outperforms the others in robustness under similar audio quality and watermark payload.

TABLE II. BIT ERROR RATES (%) UNDER TYPICAL ATTACKS

Attack	SS	ISS	CAISS	Proposed
None	1.86	0	0	0
AWGN	4.31	2.58	2.32	1.55
LPF	2.52	0.20	0.14	0.07
MP3	2.38	0.13	0.06	0.02
AAC	3.32	0.46	0.35	0.26
Requantization	3.56	0.42	0.21	0.11
Resampling	2.07	0.08	0.03	0.00

VI. Discussion and Conclusion

This paper presents a concept that the performance bound of SS-based watermarking could be extended by introducing more imperceptible MSE distortions to the host. Different from traditional embedding methods that aim at modifying the watermark, a new strategy of imperceptibly modifying the host is suggested to generate such distortions. For audio watermarking, this strategy is realized by the proposed WOHI-SS method, in which the host is inverted according to the correlation between the host and the modulated watermark sequence. It exhibits a significant improvement in watermark robustness over existing methods, while providing reasonable audio quality. Further studies will be carried out to improve the perceptual quality and extend the proposed strategy to other multimedia types such as image and video.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 61601267), by the Shandong Provincial Young and Middle-Aged Scientists Research Awards Fund of China (Grant No. BS2014DX019), and by the Youth Science Funds of Shandong Academy of Sciences (Grant No. 2014QN009).

References

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shanmoon, "Secure spread spectrum watermarking for multimedia," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[2] J. Zhong and S. Huang, "Double-sided watermark embedding and detection," IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 297–310, Sept. 2007.

[3] N. Merhav and E. Sabbag, "Optimal watermark embedding and detection strategies under limited detection resources," IEEE Transactions on Information Theory, vol. 54, no. 1, pp. 255–274, Jan. 2008.

[4] A. Valizadeh and Z. J. Wang, "Correlation-and-bit-aware spread spectrum embedding for data hiding," IEEE Transactions on Information Forensics and Security, vol. 6, no.2, pp. 267–282, Jun. 2011.

[5] P. Moulin, "Signal transmission with known-interference cancellation," IEEE Signal Processing Magazine, vol. 24, no. 1, pp. 134–136, Jan. 2007.

[6] M. Costa, "Writing on dirty paper," IEEE Transactions on Information Theory, vol. 29, no. 3, pp. 439–441, May 1983.

[7] M. L. Miller, I. J. Cox, and J. A. Bloom, "Informed embedding: exploiting image and detector information during watermark insertion," in Proceedings of IEEE International Conference on Image Processing, vol. 3, pp. 1–4, 2000.

[8] P. Zhang, Y. Li, Q. Hao, J. Jiang, and X. Chen, "High-payload spread spectrum watermarking based on informed code phase shift keying," in Proceedings of IEEE International Conference on Anti-counterfeiting, Security, and Identification, pp. 15–18, 2015.

[9] H. S. Malvar and D. A. F. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," IEEE Transactions on Signal Processing, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[10] A. Valizadeh and Z. J. Wang, "An improved multiplicative spread spectrum embedding scheme for data hiding," IEEE Transactions on Information Forensics and Security, vol. 7, no. 4, pp. 1127–1143, Aug. 2012.

[11] M. Kuribayashi, "Simplified MAP detector for binary fingerprinting code embedded by spread spectrum watermarking scheme," IEEE Transactions on Information Forensics and Security, vol. 9, no. 4, pp. 610–623, Apr. 2014.

[12] P. Zhang, S. Xu, and H. Yang, "Selective host-interference cancellation: A new informed embedding strategy for spread spectrum watermarking," IEICE Transactions on Fundamentals, vol. E95-A, no. 6, pp. 1065–1073, Jun. 2012.

[13] X. Zhang and Z. J. Wang, "Correlation-and-bit-aware multiplicative spread spectrum embedding for data hiding," in Proceedings of IEEE International Workshop on Information Forensics and Security, pp. 186–190, 2013.

[14] P. Guzik, A. Matiolanski, and A. Dziech, "Real data performance evaluation of CAISS watermarking scheme," Multimedia Tools and Applications, vol. 74, no. 12, pp. 4437–4451, 2015.

[15] Y. Xiang, I. Natgunanathan, Y. Rong, and S. Guo, "Spread Spectrum-based high embedding capacity watermarking method for audio signals," IEEE/ACM Transactions on Audio Speech and Language Processing, vol. 23, no. 12, pp. 2228–2237, Dec. 2015.