DYNAMIC STOCHASTIC RESONANCE-BASED WATERMARK EXTRACTION FROM AUDIO SIGNALS IN SVD DOMAIN

Rajib Kumar Jha, Onkar Krishna

PDPM Indian Institute of Information Technology Jabalpur, 482001, India.
(jharajib, onkarkris)@gmail.com

Kiyoharu Aizawa

The University of Tokyo, Japan.
Department of Info. & Comm. Engg.
aizawa@hal.t.u-tokyo.ac.jp

ABSTRACT

In this paper a dynamic stochastic resonance (DSR)-based watermark extraction technique from audio signal using singular value decomposition (SVD) has been presented. Watermark embedding has been done by weighted addition of the binary watermark in the singular values of the audio signal. DSR has been used in the extraction process to improve the authenticity of the extracted watermark by *utilizing* the noise or degradation introduced during different signal processing attacks. DSR is an iterative process that tunes the coefficient of possibly attacked watermarked audio signal so that effect of noise is suppressed and hidden information is enhanced. An adaptive optimization procedure has been adopted for selection of bistable parameters to achieve maximum correlation coefficient under minimum computational complexity. Resilience of this technique has been tested in presence of various signal processing attacks. Using proposed technique robust extraction of watermark is obtained without trading off the audibility of audio signal. Comparison with plain SVD-based, DWT-based and DCT-based techniques reflects that the proposed DSR-based audio watermarking scheme gives remarkably better performance in terms of correlation between original and extracted watermarks.

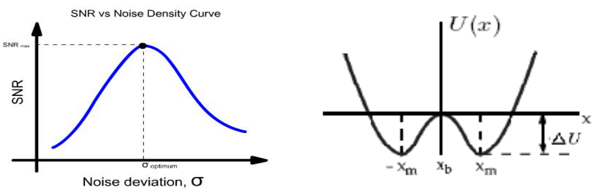
Index Terms— Audio Watermarking, Singular Value Decomposition, Dynamic Stochastic Resonance, Noise

1. INTRODUCTION

Traditionally considered a nuisance, noise has been recently found to be used to improve signal detection performance. Recent studies have convincingly shown that in non-linear systems, noise can induce more ordered regime that cause amplification of weak signals and increase the signal-to-noise ratio. This can be explained using a concept of physics stochastic resonance [1], [2], [3]. Due to widespread use of audio signals on the Internet, there a potential danger of copyright violation and Intellectual Property theft. Digital watermarking has been proposed in recent years as means of protecting multimedia contents from intellectual piracy. Digital audio watermarking is process of embedding watermarks into audio signal to show authenticity and ownership. Audio watermarking should meet the following requirements (a) Imperceptibility: the digital watermark should not affect the quality of original audio signal after it is watermarked; (b) Robustness: the embedded watermark data should not be removed or eliminated by unauthorized distributors using common signal processing operations and attacks. A significant number of watermarking techniques have been reported in recent years in order to create robust and imperceptible audio watermarks. Sun *et al.* [4] have used SR phenomenon in context of audio watermarking. A DCT-based audio watermarking method given for copyright protection of audio data by embedding watermark in high energy segments has been proposed by Dhar *et al.* [5]. An SVD-based watermarking technique has been described by [6] where the SVD of the spectrogram is modified adaptively according to the information to be watermarked. Another SVD-DWT-based watermarking technique was proposed by [7] where the effectiveness of algorithm has been brought by virtue of applying a cascade of two powerful mathematical transforms; the discrete wavelets transform (DWT) and the singular value decomposition (SVD). Most of the existing watermark technique suggests that good robustness can be achieved only at the cost of audible quality of signal. However, in this paper, a novel DSR-based technique using SVD has been proposed for robust extraction of watermark without any loss of audible quality of actual signal. DSR has been earlier used in SVD domain for improving robustness of logo extraction from images by [8]. In this paper, an analogy to Benzi's [2] double well model for global climate in the context of audio watermarking has been presented. Each of the two minima is represented by a noisy state and enhanced state of the watermarked signal respectively. The state at which signal hops into tuned state is when the correlation coefficient between original and extracted watermark is found to be maximum. In this way watermark can be extracted with good robustness. In this proposed method DSR is applied on coefficient of distorted watermarked audio signal. Our approach utilizes the noise introduced during attacks in DSR. To gauge the performance of the watermark embedding by measuring imperceptibility of the watermark, metrics Signal-to-noise ratio (SNR) has been calculated. Its value should be greater than 13 dB for good audibility. Another objective measure, *Mean Opinion Score (MOS)* which is the average score (between 1 and 5) of audio quality of a signal given by 5 different listeners. To measure the similarity between original and extracted watermark, Correlation coefficient, ρ has been computed. The correlation factor ρ may take values between 0 (random relationship) to 1 (perfect linear relationship).

2. DYNAMIC STOCHASTIC RESONANCE

In order to exhibit SR, a system should possess three basic properties: *a non-linearity in terms of threshold*, *a subthreshold signal* like a signal with small amplitude, and *a source of additive noise*. This phenomenon occurs frequently in bistable systems or in systems with threshold-like behavior. The general behavior of SR mechanism shows that at lower noise intensities the weak signal is unable to cross the threshold, thus giving a very low SNR. For large noise intensities the output is dominated by the noise, also leading to a low SNR. But, for moderate noise intensities, the noise allows the signal to cross the threshold giving maximum SNR at some optimum noise level. Thus, a plot of SNR as a function of noise intensity shows a peak at an optimum noise level as shown in Fig. 1(a). A classic one-dimensional non-linear dynamic



(a) Signal-to-noise ratio vs. noise density

(b) Bistable double well potential system

Fig. 1. SR in double well potential

system that exhibits stochastic resonance is modeled with the help of Langevin equation of motion [9]. Addition of a periodic input signal $[B\sin(\omega t)]$ to the bistable system makes it time-dependent whose dynamics are governed by Eq. 1.

$$\frac{dx(t)}{dt} = -\frac{dU(x)}{dx} + B\sin(\omega t) + \sqrt{D}\xi(t) \qquad (1)$$

where $U(x)$ is a bistable quartic potential (Fig. 1(b)) given in Eq. 2.

$$U(x) = -a\frac{x^2}{2} + b\frac{x^4}{4} \qquad (2)$$

Here, a and b are positive bistable double-well parameters. The double-well system is stable at $x_m = \pm\sqrt{\frac{a}{b}}$ separated by a barrier of height $\Delta U = \frac{a^2}{4b}$ when the $\xi(t)$ is zero. Here B and ω are the amplitude and frequency of the periodic signal respectively and there an additive stochastic fluctualtion (noise) $\xi(t)$ with intensity D. It is assumed that the signal amplitude is small enough so that in the absence of noise it is insufficient to force a particle to move from one well to another. It is assumed that the signal amplitude is small enough so that in the absence of noise it is insufficient to force a particle of unit mass to move from one well to another. It, therefore, fluctuates around its local stable states. When a weak periodic force and noise are applied to the unit mass particle in the potential well, noise-driven switching between the potential wells takes place only at some 'resonant' values of noise. This noise-induced hopping is synchronized with the average waiting time, between two noise-driven inter-well transitions that satisfies the time-scale matching between signal frequency and the residence times of the particle in each well. Maximum SNR is achieved when $a = 2\sigma^2$ by differentiation of SNR expression [8]. Thus SNR has maximum value at an intrinsic parameter, a, of the dynamic double well system. The other parameter b can be obtained using parameter a. For weak input signal, condition $b < 4\frac{a^3}{27}$ is required to ensure subthreshold condition [8]. Solving the stochastic differential equation given in Eq. 1 using the stochastic version of Euler-Maruyama's iterative discretized method as follows [10].

$$x(n+1) = x(n) + \Delta t[ax(n) - bx^3(n) + Input] \qquad (3)$$

where $Input = Bsin(\omega(t) + D\xi(t)$ denotes the sequence of input signal and noise. In this paper, $Input$ can be replaced by audio signal as they are considered to contain information both related to signal as well as noise due to attacks [8]. Here Δt is the sampling time, taken as 0.015 experimentally. In case

3. WATERMARK EMBEDDING

The cover signal is an audio signal while the watermark is an image, the reason behind taking logo image (2-D signal) as a watermark is to measure robustness by visualizing visual extracted logo. The following steps have been followed for embedding the watermark into the audio signal [6].

Let $X = \{x(i), 1 \leq i \leq L\}$ represent a host audio signal of length $L = p^2$. $W = w(i,j)$, such that $1 \leq i \leq p, 1 \leq j \leq p$ is a binary logo of size $p \times p$, to be embedded in host audio signal. Let $A = (A_{i \times j})_{p \times p}$ be a matrix representation of host audio signal, we represent signal as a matrix mathematical simplicity. X, with SVD of the form $A = UDV^T$, where U and V are orthogonal matrices. The columns of U are called left singular vectors while those of V are right singular vectors. D is diagonal matrix with nonnegative elements, called the singular values. Let $u \leq p$ be the rank of matrix A.

The nonzero elements $\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_u$ of D are the singular values of matrix A. The SVD of the reshaped cover signal matrix provides a medium to embed a 2D watermark pattern directly. In order to ensure the inaudibility (to guarantee that the modifications are below the HAS hearing level) the embedding watermark message is shaped with singular values of original/host audio signal, thus the embedding watermark is modified adaptively with embedded coefficients.

Step 1: The audio signal X of length p^2 is converted into a matrix $A = (A_{i \times j})_{p \times p}$, having $p \times p$ elements.

Step 2: Apply singular value decomposition on host audio signal matrix A producing three matrices as U, D, V. The watermark $W = w(i,j)$, $p \times p$ is added to the diagonal matrix D with scaling factor α as follows.

$$w_d(i,j) = \lambda_i + \alpha \lambda_i w(i,j) \tag{4}$$

where α is the watermark amplification factor, taken as 0.02 and w_d is the modified watermark. SVD of this modified watermark is performed as $W_d = U_w D_w V_w^T$.

Step 3: The singular matrix of this decomposition, D_w is used with singular vectors of cover signal to reconstruct the watermarked audio signal as $A_w = U D_w V^T$. SNR of this watermarked audio signal with respect to the original cover audio signal is computed to assess the audible quality. Perceptual audio quality measure is also computed using mean opinion scores given by 5 listeners.

4. PROPOSED DYNAMIC STOCHASTIC RESONANCE-BASED WATERMARK EXTRACTION

The watermarked audio signal may be subjected to intentional or unintentional attacks during transmission over a channel. We used U_w, V_w and D as a key during extraction process. We transmitted these key matrics through secured channel to receiver. The proposed DSR-based extraction process follows the steps as given under:

Step 1: Apply DSR on the possibly attacked audio signal, $A_w^{'}$ as follows.

(a) Initialize $A_{dsr}(0)$=0, $a = 2{\sigma_0}^2$, $b = 0.005 \times 4a^3/27$. Value of bistable system a is initialized following the condition of maximizing SNR for a DSR system, where σ_0 is standard deviation of noise approximated as standard deviation of attacked, watermarked audio signal. while m is a number less than 1 to ensure sub threshold condition of signal.

(b) Using iterative equation given in Eq. 5, compute tuned audio signal matrix. Here $A_{dsr}(n+1)$ is the DSR enhanced (tuned) audio signal, n is the iteration number; a, b and Δt are the bistable system parameters.

$$A_{dsr}(n+1) = A_{dsr}(n) + \Delta t[a A_{dsr}(n) - b {A_{dsr}}^3(n) + A_w^{'}] \tag{5}$$

Step 3: Perform SVD on DSR-tuned audio signal $A_{dsr}((n+1)$ to obtain $U^{'}, D_w^{'}$ and $V_T^{'}$. Using this $D_w^{'}$ and singular vectors U_w and V_w obtained in **Step 2** of embedding process, find the tuned modified matrix $W_d^{'} = U_w D_w^{'} V_w^T$.

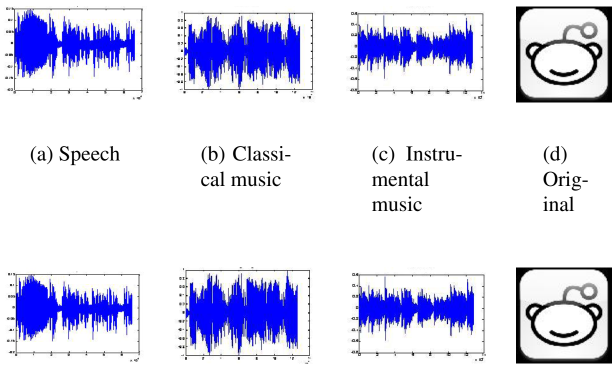
Step 4: Extract the watermark bits from this matrix using

$$w^{'}(i,j,n+1) = \frac{1}{\alpha}\left(\frac{w_d^{'}(i,j)}{\lambda_i} - 1\right) \tag{6}$$

Step 5: Calculate cross-correlation coefficient $\rho(n+1)$ between extracted watermark and original watermark after each iteration. To make the algorithm adaptive, repeat **Step 2** to **Step 5** for increasing number of iteration n and get the value n for which $\rho(n)$ becomes maximum.

5. SIMULATION RESULTS AND OBSERVATIONS

Three cover audio signals (Speech, Classical music and Instrumental music) that have been tested for this technique are shown in Fig. 2(a), Fig. 2(b) and Fig. 2(c) respectively. Watermarked signal for these cover audio signals have been shown in Fig. 2(e), Fig. 2(f) and Fig. 2(e) respectively. Fig. 2(d) shows the original binary watermark and Fig. 2(h) shows extracted watermark in the absence of any attacks from Speech signal. Assumed parameters m=0.005 and Δt=0.5.



(a) Speech (b) Classical music (c) Instrumental music (d) Original

(e) Speech, SNR=24.69 (f) Classical Music, SNR=33.61 (g) Instrumental Music, SNR=30.82 (h) Extracted

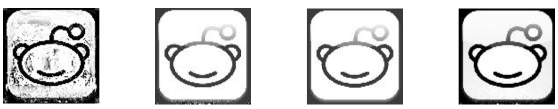
Fig. 2. Top fig is original cover audio signal and below is watermarked signal

5.1. Quality of Watermarked Audio Signal

Inserting watermark in audio signal introduce a small amount of distortion (noise) in audio signal. Quality of host signal after adding watermark is calculated using parameter known as SNR as discussed in Section 1. SNR values for different values of α=0.03 (as shown in Fig. 2 suggest that the audibility of the watermarked audio signals is not impaired by the watermark embedding process. The MOS values for all watermarked signals are, Classical - 3, Speech - 4 and Instrumental - 4, indicating that good imperceptibility is obtained for both the cover signals.

5.2. Quality of Extracted Watermarks

With α=0.03, in case of speech signal, correlation of extracted watermark with original watermark without any attack is 1.0000, in both technique. When tested for watermarked audio corrupted by different type of attacks the DSR based technique reaches very high correlation as shown in Fig. 3. The plain frequency-based techniques are dependent on value of α for robustness. Response under various signal processing attacks such as adding dynamic noise ($add - dyn - noise$), amplification (AMP), zero crossing (ZC), compression ($Compress$), gaussian noise ($Gaussian$), extrastereo, invert, voice removal, smoothing ($Smooth1$ and $Smooth2$) and echo ($Echo$) [11] for the three cover audio signals have been tested and tabulated in Table 1. Correlation coefficient values obtained for three cover audio signals in comparison with existing frequency domain audio watermarking techniques [6], [12], [5] have been tabulated in Table 2. It is therefore observed that for good audio quality of watermarked audio signal, remarkably better quality of extracted watermark is obtained using proposed DSR-based technique than that obtained using plain SVD, DCT or DWT watermarking. Graph showing correlation coefficient, $\rho(n)$ with respect to iteration (n) has been shown in Fig. 4. At an average for amplification attack, the peak correlation coefficient is obtained at n=3 iterations.



(a) Noise (b) AWGN (c) Zero Cross. (d) Amp.

(e) Cut sample (f) Echo (g) Smooth1 (h) Smooth2

Fig. 3. Extracted watermarks for various attacks by proposed SVD and DSR-based technique

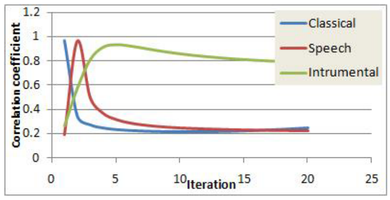
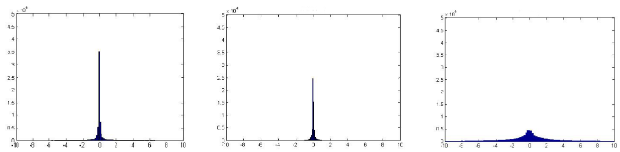


Fig. 4. Number of iteration (n) vs. correlation coefficient (ρ) for amplification attack

5.3. Discussion

The basic mechanism of DSR for improvement of robustness is attributed to the way DSR modify the distribution of the attacked audio signal values. It is observed that attack (echo) shrinks the distribution of audio signal values making accurate extraction of logo difficult (Fig. 5). However, after DSR iteration $n = 6$, the distribution is flattened increasing overall energy of the audio signal. This increases the energy of hidden data also, making accurate extraction of watermark easier.



(a) Watermarked speech (b) Echo attack (c) DSR iteration n=6

Fig. 5. Distribution of SVD coefficients of audio signal

6. CONCLUSION

An adaptive DSR-based algorithm for watermark extraction using SVD has been proposed and analyzed in this paper. The most striking feature of the proposed technique is the remarkable improvement in robustness of watermark extraction. Thus, authenticity is increased without compromising with audibility of audio signal. Improvement in robustness is achieved due to rearrangement of distribution of distorted signal coefficients and transferring them from a weak state to a strong state by an iterative DSR procedure. The noise introduced during attack itself is utilized in DSR iteration to suppress the effect of noise in watermark extraction. An adaptive algorithm is used to ensure minimum computational complexity. Robustness has been evaluated against different signal processing attacks and results suggest that our proposed technique gives better robustness than the plain SVD, DCT and DWT-based techniques.

Table 1. Correlation coefficient values obtained from the proposed DSR-based technique for three cover audio signals- speech, classical music and instrumental music for various attacks

	Speech		Classical		Instrumental	
Attacks	**SVD-DSR**	SVD	**SVD-DSR**	SVD	**SVD-DSR**	SVD
AMP	**0.9977**	0.9620	**0.9977**	0.9620	**0.9977**	0.9620
ZC	**0.9937**	0.8326	**0.9262**	0.1483	**0.9835**	0.9630
Gaussian	**0.9828**	0.9784	**0.9975**	0.9839	**0.9998**	0.9995
Echo	**0.8235**	.0278	**0.8391**	0.2259	**0.4260**	0.9371
Extrastereo	**0.9228**	0.9174	**0.9188**	0.9174	**0.9188**	0.9174
Invert	**1.0000**	1.0000	**1.0000**	1.0000	**1.0000**	1.0000
Smooth1	**0.9667**	0.8527	**0.9955**	0.9955	**0.9940**	0.9926
Stat1	**0.9547**	0.8181	**0.9821**	0.9598	**0.9856**	0.9573
Voice removal	**1.0000**	1.0000	**1.0000**	1.0000	**1.0000**	1.0000
Add-dyn-noise	**0.8613**	0.5216	**0.9288**	0.8562	**0.9295**	0.7747
Cutsample	**0.8124**	0.4731	**0.8645**	0.5794	**0.8942**	0.5120

7. REFERENCES

[1] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni, “Stochastic resonance,” *Reviews of Modern Physics*, vol. 70, pp. 223–270, 1998.

[2] R. Benzi, A. Sutera, and A. Vulpiani, “The mechanism of stochastic resonance,” *Journal of Physics A*, vol. 14, pp. L453–L457, 1981.

[3] B. McNamara and K. Wiesenfeld, “Theory of stochastic resonance,” *Physical Review A*, vol. 39, no. 9, pp. 4854–4869, 1989.

[4] Shuifa Sun, Jiang Ming, and Liu Xing-qiong, “A blind audio watermarking based on stochastic resonance signal processor,” in *Proc. IEEE International Conference on Signal Processing*, 2008, pp. 2185–2188.

[5] P. K. Dhar, M. I. Khan, and S. Ahmad, “A new dct-based watermarking method for copyright protection of digital audio,” *International Journal of Computer Science and Information Technology*, vol. 2, no. 5, pp. 91–101, 2010.

[6] Hamza Ozer, Blent Sankur, and Nasir Memon, “An svd-based audio watermarking technique,” in *Proc. of the 7th Workshop on Multimedia and Security*. 2005, ACM New York, NY, USA.

[7] Ahmad Mohammad Ali Al-Haj, Christina Twal, “Hybrid dwt-svd audio watermarking,” .

[8] Rajlaxmi Chouhan, Rajib Kumar Jha, Apoorv Chaturvedi, Toshihiko Yamasaki, and Kiyoharu Aizawa, “Robust watermark extraction using svd-based dynamic stochastic resonance,” in *Proc. IEEE International Conference on Image Processing*, 2011, vol. 11, pp. 2801–2804.

[9] H. Risken, *The Fokkar Plank Equation*, Berlin: Springer verlag, 1984.

[10] T. C. Gard, *Introduction to Stochastic Differential Equations*, New York: Marcel-Dekker, 1998.

[11] Christian Kraetzer Jana Dittmann, *Audio Benchmarking Tools and Steganalysis*, Katholieke Universiteit Leuven (KUL), 2006.

[12] Ali Al-Haj, Ahmad Mohammad, and Lama Bata, “Dwt-based audio watermarking,” *The International Journal of Information Technology*, vol. 8, no. 3, pp. 326–333, 2011.

Table 2. Correlation coefficient values obtained from the proposed DSR-based technique for Speech Signal in comparison with plain SVD-based [6], plain DWT-based [12] and plain DCT-based [5] techniques for various attacks

Attacks	**SVD-DSR**	SVD	DWT	DCT
AMP	**0.9977**	0.9620	0.9467	0.9353
ZC	**0.9937**	0.8326	0.9141	0.9042
Gaussian	**0.9828**	0.9784	0.9808	0.9672
Echo	**0.8235**	0.0278	0.7434	0.7202
Extrastereo	**0.9228**	.09174	0.5762	0.4932
Invert	**1.0000**	1.0000	-0.0625	-0.0203
Smooth1	**0.9667**	0.8527	0.9610	0.0.9135
Stat1	**0.9547**	0.8181	0.9857	0.9262
Voice removal	**1.0000**	1.0000	1.0000	1.0000
Add-dyn-noise	**0.8613**	0.5216	0.6250	0.5736
Cutsample	**0.8012**	0.4731	0.7527	0.7322