Acta Cybernetica — *online–first paper version* — pages 1–19.

Semi Fragile Audio Crypto-Watermarking based on Sparse Sampling with Partially Decomposed Haar Matrix Structure

Electa Alice Jayarani Appadurai[a], Mahabaleswara Ram Bhatt[b], and Geetha D.D.[c]

Abstract

In the recent era the growth of technology is tremendous and at the same time, the misuse of the technology is also increasing with an equal scale. Thus, the owners have to protect the multimedia data from the malicious and piracy. This has led the researchers to the new era of cryptography and watermarking. In the traditional security algorithm for the audio, the algorithm is implemented on the digital data after the traditional analog to digital conversion. But in this article, we propose the crypto–watermarking algorithm based on sparse sampling to be implemented during the analog to digital conversion process only. The watermark is generated by exploiting the structure of Haar transform. The performance of the algorithm is tested on various audio signals and the obtained SNR is greater than 30dB and the algorithm results in good robustness against various signal attacks such as echo addition, noise addition, reverberation etc.

Keywords: audio, watermarking, cryptography, compressive sensing

1 Introduction

The most common and widely used security algorithm for the multimedia files is digital algorithms. The multimedia data can be the image, audio, video, text, etc. Mainly there are two ways to achieve the privacy in digital data, namely, watermarking and cryptography [11]. The digital watermarking is defined as embedding the highly decryptable watermark into the digital data without harming the content of the original host signal. Whereas in cryptography the data would be in

[a]Research Scholar, Department of Electronics and Communication, Reva University, Bangalore, India. E-mail: electalice@gmail.com, ORCID: https://orcid.org/0000-0002-5117-6917.

[b]Professor, Department of Medical Electronics, BMS College of Engineering, Bangalore, India. E-mail: bhatt.mr@rediffmail.com, ORCID: https://orcid.org/0000-0002-6921-036X.

[c]Professor, Department of Electronics and Communication, Reva University, Bangalore, India. E-mail: dgeetha@reva.edu.in, ORCID: http://orcid.org/0000-0002-7788-5615.

DOI: 10.14232/actacyb.280899

disguise form to protect its content. In other words, Cryptography converts the intelligible data into unintelligible data which appears as meaningless for attackers. By seeing the data one can tell the data is encrypted but cannot decrypt without the proper secret key. If the data is decrypted the data is no longer protected. Both the algorithm should maintain the robustness nature to protect the secret message. On the other hand, the privacy in watermarking is not strictly inevitable but in cryptography, it has to be private by definition. For example, the watermark presence on the rupee note can be easily seen by everyone against the light.

In this article, we propose the algorithm to protect the audio signals from the piracy. As the human Auditory System (HAS) is more sensitive than the Human Visual System (HVS) [11], the audio watermarking becomes a very tedious task. The audio data security has been under research for many years but still, it is falling short of safety requirements and it is vulnerable to attack, privacy and piracy. The natural audio signal that is audible by the Human ear originates from acoustic variation. These acoustic signals are converted to analog and subsequently digital data using Shannon sampling theory. The encrypted key or watermarking is carried out on the obtained digital data for protection. A large amount of research in watermarking is centered on digital techniques which are more prone to attack as shown in Fig. 1.

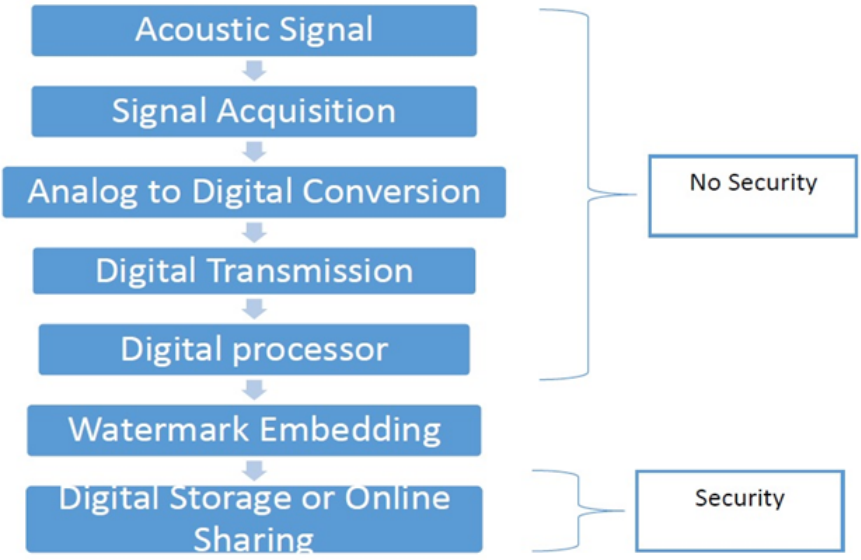

Figure 1: Existing Methods Flowchart

To overcome this problem, in this paper, embedding the crypto-watermark signature on the audio during the time of digital conversion as shown in Fig. 2 is studied and experimented.

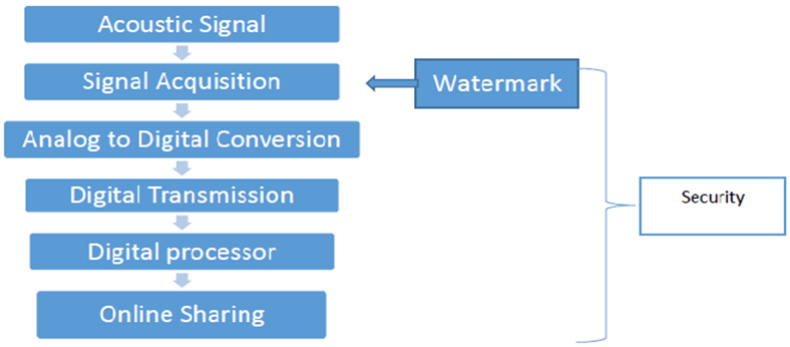

Figure 2: Proposed Algorithm Flowchart

In the past decade, there has been a paradigm shift in approaches to signal acquisition that explores and employs sparse coding or compressing sensing [3, 12, 13, 14]. By compressive sensing the audio with the watermark, the data is referred as 'digital information data' instead of typical digital audio data, which precludes from direct conversion to analog audio unless the audio can be recovered using mathematical programming techniques only. The advent of this technique is to embed the watermark with the secret key at the time of digital to analog conversion without altering the perceptual quality of the audio signal. By using only, the mathematical programming technique the audio can be converted and can be played using a transducer.

2 Existing methods

In the past years, various research had been undergone to protect the ownership of audio files and a various algorithm is developed based on Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), etc. In [6], Guo et al. propose a transform domain watermarking algorithm. By altering the DCT coefficient the watermark is embedded into the host and the algorithms average Signal to Noise Ratio (SNR) reaches up to 20dB. A novel audio watermarking algorithm based on the randon transformnumber and DWT was defined by Cairong Li et al. [10]. A new adaptive audio watermarking algorithm based on Empirical Mode Decomposition is introduced by Khaldi and Boudraa [9] and the average SNR reaches up to 25dB. In [1], the author attempts to implement a baseline audio watermarking system that embeds the information by modulating the phase in Weighted Overlap-Add Algorithm (WOLA). The algorithm gives SNR values from 0 to 25dB. In [7] blind audio watermarking is proposed based on a combination of Discrete Wavelet Packet Transformation (DWPT), Singular Value Decomposition (SVD) and Quantization Index modulation (QIM). The

author Fallahpour and Megias [4] venture an innovative method of embedding the audio watermark. The Fibonacci series is used to select the FFT samples of the host signal to embed the watermark. In all the methods the acoustic signal is converted to digital data using traditional analog to digital conversion (ADC) and the algorithms are implemented on the digital data.

3 Block diagram

The general digital audio watermarking process is shown in Fig. 3. From the performer through the microphone the audio signal is transmitted to the processor where the signal is converted into digital and watermark embedding is done. The watermarked data can be transmitted or can be stored digitally. At the receiver side, the signal is converted into audio and played through the speaker. Thus, the algorithms cannot be used for the live audio concert.

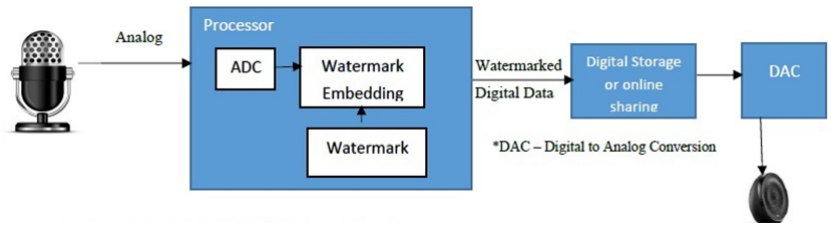

Figure 3: General Digital Audio Watermarking process

To overcome this problem in this paper we propose a compressive sensing based crypto–watermarking algorithm to be implemented during the process of ADC only. The general block diagram of the compressive sensing based crypto-watermarking algorithm for audio is shown in Fig. 4.

Here we propose a customized microphone where the watermark is embedded in the time of signal acquisition and the watermarked digital data can be transmitted or can be stored digitally. The analog data can be recovered only by the customized speaker where the security key and watermark are embedded. The traditional speaker cannot retrieve the data. The customized microphone and speaker block diagram are shown in Fig. 5.

4 Compressive sensing and its role for audio security

Essentially, the compressive sampling (CS) is a method of converting the analog signal into a digital information with sparse. This non-uniform sampling yields fewer sample data, which can be used to recover the signal using a mathematical

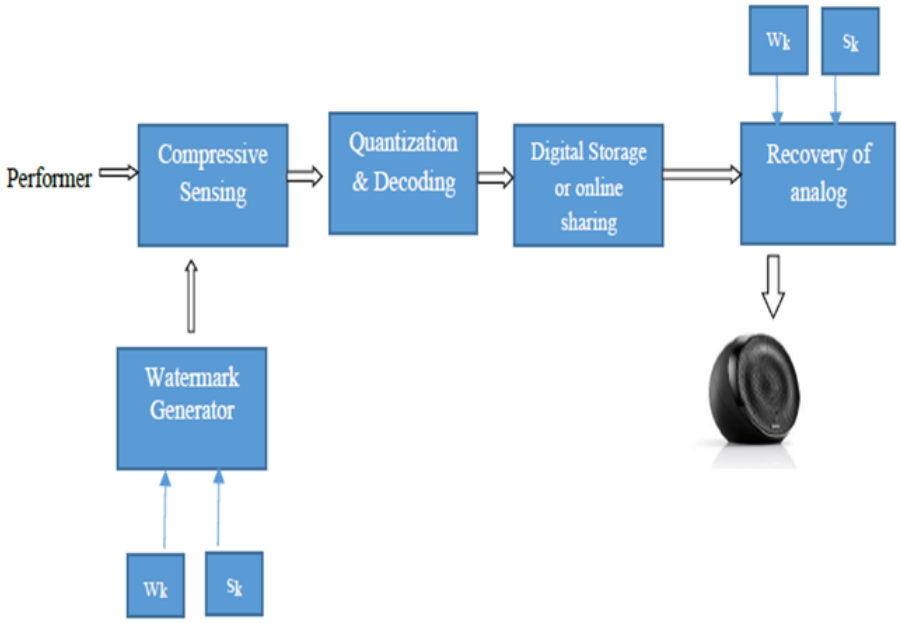

Figure 4: Crypto–Watermarking Block Diagram

convex programming. This is in contrast to the conventional analog to digital conversion technique which exploits digital filtering technique based on Shannon uniform sampling principle.

Let $x \in R^n$ be a one dimensional (1-D) original audio signal and the signal is considered as K-sparse or K non-zero entries. The transform matrix vector representation with the orthonormal basis matrix $\Psi \in R^{n \times n}$ is $X = \Psi x$ with x is a K-sparse signal.

The method of obtaining linear measurement data vector $y \in R^m$ from an incoherent sampling or sensing matrix $\phi \in R^{mxn}$ (m $\ll$ n) is expressed as $y = \phi \Psi x$.

On denoting matrix $\Theta = \phi \Psi$ as compressive sensing process we get

$$y = \Theta x \ . \tag{1}$$

By finding solutions to an underdetermined linear system of equation (1), the original signal can be reconstructed. In underdetermined linear system, the system has infinite number of solutions and more unknowns than the equations. Most common methods to solve the sparse approximation are Basis Pursuit and Orthogonal Matching Pursuit methods. In basis Pursuit method, the sparse approximation problem can be replaced as convex problem, hence the same is used for the recovery in the proposed method.

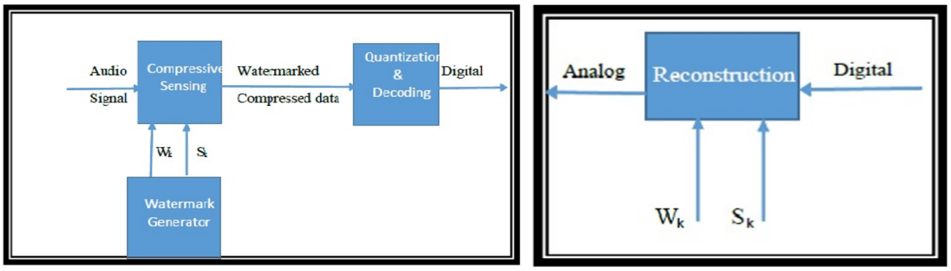

Figure 5: (a) Customized Microphone Block Diagram (left) (b) Customized Speaker (right)

The sparse problem in Basis Pursuit is given as

$$\min(\|x\|_0) \text{ subject to } y = \Theta x \ , \tag{2}$$

where $y \in R^m$ is the measured vector, ϕ is the $m \times n$ matrix and $x \in R^n$ is the vector to be recovered. In the equation (2), the norm-0, $||.||_0$ is non-convex and difficult to solve. It is an NP-hard (Non-deterministic Polynomial-time hardness) problem. Therefore, it is replaced with l_1-norm and it is given as

$$\min(\|x\|_1) \text{ subject to } y = \Theta x \ . \tag{3}$$

It can be recast as Linear Programming problem (LP) and is given as

$$\begin{aligned} \min f^T x \text{ subject to } & y = \Theta x \\ & x \geq 0 \ , \end{aligned}$$

where $f^T x$ is the objective function, $y = \Theta x$ is collection of equality constraint and $x \geq 0$ is set of bounds. By adding new variable, the nonlinearity is recast to the set of constraints and it is given as

$$\begin{aligned} \min \sum_{i-1}^{n} U_i \text{ subject to } & -u \leq x \leq u \\ & y = \Theta x \end{aligned}$$

Or it can be written as

$$\begin{aligned} \min \sum_{i-1}^{n} U_i \text{ subject to } & -x_i - u_i \leq 0, \ i = 1, 2, \ldots, n \\ & x_i - u_i \leq 0, i = 1, 2, \ldots, n \\ & y = \Theta x \end{aligned} \tag{4}$$

There are many algorithms to solve the basis pursuit problem such as simplex method and primal–dual interior point method. For high accuracy, the primal dual method is used with Newton method combined with modified KKT (Karush-Kuhn-Tucker) condition for search criteria.

For example, consider $n = 2$,

$$f_{u1} = x_1 - u_1$$
$$f_{u2} = -x_1 - u_1$$

And the corresponding dual variable is considered as λ_1 and λ_2 and given as

$$\lambda_1 = -\frac{1}{f_{u1}}$$
$$\lambda_2 = -\frac{1}{f_{u2}}$$

The modified KKT condition for the residual $r_t = (x, \lambda, \upsilon)$ is given as

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \Theta^T \upsilon = 0$$
$$-\lambda_i f_i(x) = \frac{1}{t} \quad i = 1, 2, \ldots, m$$
$$\Theta x = y$$

For $t > 0$, it is given as

$$r_t(x, \lambda, \upsilon) = \begin{pmatrix} \nabla f_0(x) + Df(x)^T \lambda + \Theta^T \upsilon \\ -\mathrm{diag}(\lambda) f(x) - \frac{1}{\tau} 1 \\ \Theta x - y \end{pmatrix} \tag{5}$$

where $f : R^n \to R^m$ and the matrix Df is its derivative

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} \text{ and } Df(x) = \begin{pmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{pmatrix}.$$

If x, λ, υ and $r_t(x, \lambda, \upsilon) = 0$, then $x = x^*(t)$, $\lambda = \lambda^*(t)$ and $\upsilon = {}^*(t)$. x is primal feasible, and λ, ν are dual feasible. The duality gap is $\tau = \frac{m}{t}$.

The first term of equation (5) is called dual residual, 2nd term is called centrality residual and 3rd term is primal residual. For a fixed time t, at a point (x, λ, υ) that satisfies $f(x) < 0, \lambda > 0$ the Newton's step is used to solve $r_t(x, \lambda, \upsilon) = 0$.

$$y = (x, \lambda, \upsilon), \ \triangle y = (\triangle x, \triangle \lambda, \triangle \upsilon)$$

$$\begin{pmatrix} \nabla^2 f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x) & Df(x)^T & \Theta^T \upsilon \\ -diag(\lambda) Df(x) & -diag(f(x)) & 0 \\ \Theta & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta \upsilon \end{pmatrix} = - \begin{pmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{pmatrix} \tag{6}$$

The solution of equation (6) will be the primal dual search direction. For the primal-dual interior point method, we use the surrogate duality gap. For any x that satisfies $f(x) < 0$, $\lambda \geq 0$ it is defined as $\eta(x, \lambda) = -f(x)^T\lambda$.

If x is a primal feasible, and λ, υ are dual feasible, which means $r_{pri} = 0$ and $r_{dual} = 0$ then the surrogate gap will be the duality gap. In general, the steps to compute the optimal solution is as follows. The inputs are a point x that satisfies $f(x) < 0, \lambda > 0, \mu > 1\varepsilon_{feas} > 0, \varepsilon > 0$.

1. Set $t = \frac{\mu m}{\eta}$.
2. Compute primal dual search direction using equation (6).
3. We determine the step length $s > 0$ and compute $y = y + s\Delta y$ until $\|r_{pri}\|_2 \leq \varepsilon_{feas}$, $\|r_{dual}\|_2 \leq \varepsilon_{feas}$, and $\eta \leq \varepsilon$.
4. For the implementation, the step length is chosen in the range of $0 < s \leq 1$. The step length tracking is started with $s = 0.99 \cdot \min\{1, \min\frac{-\lambda_i}{\triangle\lambda_i} \mid \triangle\lambda_i < 0\}$, $i = 1, 2, \ldots, m$. Multiply the s by $\beta \in (0, 1)$ until we have $\|r_\tau(x + s \triangle x, \lambda + s \triangle \lambda, \upsilon + \triangle\upsilon\|_2 \leq (1 - \alpha s).\|r_\tau(x, \lambda, \upsilon)\|_2$ where α is set as 0.01.
5. Continue the steps until the optimal value of x is found.

5 Haar transform and its orthogonal property

Haar Transform is the simplest and the fastest wavelet transform. The Haar function is denoted as $h_k(x)$ and will fall in the closed interval of $[0, 1]$. Whereas the k is the order of the function and it is decomposed into two parameter such as $k = 2^p + q - 1$, $k = 0, 1, \ldots, N - 1$ where $N = 2^n$, $0 \leq p \leq n - 1$, $0 \leq q \leq 2^p$.

The Haar function is defined as

$$h_0(x) \equiv h_{00}(x) = \frac{1}{\sqrt{N}}, \quad x \in [0, 1]$$

and

$$h_k(x) \equiv h_{pq}(x) = \frac{1}{\sqrt{N}} \begin{cases} 2^{\frac{p}{z}} \frac{q-1}{2^p} \leq x < \frac{q-0.5}{2^p} \\ -2^{\frac{p}{z}} \frac{q-0.5}{2^p} \leq x < \frac{q}{2^p} \\ 0 \quad \text{otherwise.} \end{cases}$$

The amplitude and the width of the function which involves the value other than zero is given by p and position of the non-zero value is given by q. The Haar transform matrix for the $N = 2$ is given below:

$$H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}.$$

It is observed that $H = H^*$ and $H^{-1} = H^T$ therefore $H^T H = I$ where I is the identity matrix.

6 Privacy preserving crypto-watermarking technique

Typically, the intent of both cryptology and watermarking is to add a signature into the data to make it secure from an unintended audience and to maintain privacy and authenticity while communicating through the unsecured channels with robustness to attacks. But the significant difference is that in cryptology, both data and signature are invisible, whereas in watermarking the data could be visible but signature may or may not be visible. The current exploration has both the features which we refer to as crypto-watermarking technique.

For our algorithm, we create a matrix U which can be a unitary matrix or permutation matrix since both has very interesting properties. For example, let's consider U as a unitary matrix and considering the property of unitary matrix

$$UU^T = U^T U = I \ . \tag{7}$$

Applying the equation (7) to (1) we get

$$y = \Theta x = \phi \Psi x = (U\phi^T)^T (U\Psi)x \tag{8}$$

or

$$y = \Theta x = \phi \Psi x = (U^T\phi^T)^T (U^T\Psi)x \ \ . \tag{9}$$

Note here the matrix x is the segmented audio frame of the original host signal. Based on the above relationship we now formulate the sensing matrix and the transform matrix by using either equation (8) or (9). By calculating the scaling factor, the equation (7) can be rewritten as

$$UU^T = U^T U = \frac{1}{n} I \ \ . \tag{10}$$

Therefore, we can view y as

$$y = \Theta x = \phi \Psi x = (U\phi^T)^T (U\Psi)x$$

$$y = \Theta x = \phi \Psi x = (U^T\phi^T)^T (U^T\Psi)x \ \ .$$

7 Proposed algorithm

Generation of K-sparse signal

The original host signal is divided into frames and the input audio sequence from an audio frame is $x \in R^n$ (e.g., Figure 6) with K sparse.

Process of generating a watermark signature

1. Consider U=H, a Haar matrix.

2. Form $H = Q_1Q_2Q_3 \cdots Q_jR$ where Q_j is an orthogonal matrix and R is an upper triangular matrix which in turn is non-orthogonal matrix.

3. Perform the various signal function on Q to generate a watermark and is given as $W = \text{signalfunction}_i(Q)$ where the signal function can be circular shift, addition, etc. on the decomposed orthogonal matrix without affecting the orthogonal property. The signal function and "i" times is considered as extra security key (S_k).

4. The generated watermark is considered as watermark key (W_k).

Process of embedding watermark signature in compressive sensed data

Let us consider equation (8).

1. Decompose the Haar matrix and generate the watermark key and secret key.

2. Obtain the shuffled audio matrix as $X = (U\Psi)x$.

3. Obtain $A = (U\phi^T)^T$.

4. Obtain watermarked data matrix as $Y = AX$.

Process of Recovery of Signal from compressive signal

In order to recover the signal from equation (1), the primal dual interior method is used. The recovery algorithm explained in section 4 is implemented in MATLAB and the signal is recovered. Depends on the length of the audio signal, the number of iterations varies. Table 1 lists the number of iterations for the different audio files.

Table 1: Number of iterations

Audio file	**Number of iterations**	**Duration(sec)**
Guitar	9	16.52
Flute	14	37.47
Bass	21	46.53

8 Results and discussion

In this section, we concentrate on the audio quality aspects arising due to compressed sensing that exploits various k-sparse audio data. Subsequently, we demonstrate and highlight a few experimental results of the proposed semi-fragile audio crypto-watermarking based on compressed sensing while acquiring audio clips and also audio data recovery processes. The experiment involves schemes such as crypto-watermarking signature generation, embedding the watermark signature and l_1 recovery algorithm for the recovery of the signal. And we have compared the quality assessment of the audio recovery using the proposed algorithm with and without watermarking signatures. The proposed algorithm is implemented using MATLAB 2016 in Intel Core i5 processor.

Generating K-sparse data for experimentation

A set of 10 source audio clips are chosen for the experiment. All the clips are mono-channel with less than 60 seconds duration sampled with 44.1 KHz having audio data width as 8 bits. All the audio clips generated includes solo musical instruments like violin, guitar, piano, flute, equinox, bass, Handel, track, Mary Song, Backstreet boys song, Crazy Frog - Axel F, Emilie big world, and different frequency clips. Table 2 lists the different audio clips names, duration, length and the sampling frequency.

Table 2: Experimented audio file's details

Audio	**Length**	**Duration**	**Sampling Frequency (Hz)**
bass	525200	11 s	44100
Guitar	90309	2 s	44100
Piano	409101	9 s	44100
Handel	73113	8s	8192
violin	305172	6s	44100
flute	346724	7s	44100
tone	384000	8s	48000
Mary	319725	7s	44100
Backstreet boys	1323000	30s	44100
Emilie big world	1323000	30s	44100
Irish Whistel	1323000	30s	44100
100Hz	220500	5s	44100
250Hz	220500	5s	44100
440Hz	220500	5s	44100
1KHz	220500	5s	44100

For better implementation, the source signal is reduced to frames with the samples of 256 for each frame. Many natural signals are pithy when it is expressed in an appropriate basis. The example is shown in below Fig. 6(a) of the source signal and its transform in Fig. 6(b).

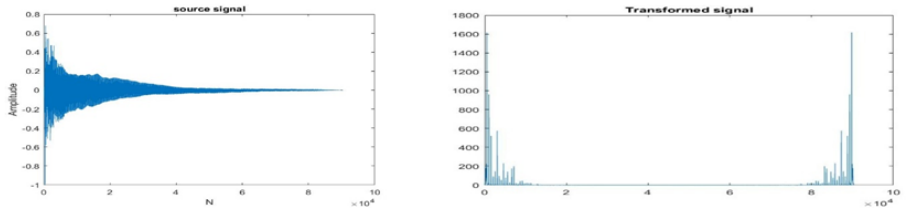

Figure 6: (a) Source Signal (b) Transformed signal

Based on observation, it is evident that the most coefficients are very small and negligible and at the same time only a few coefficients would comprise of a significant amount of information. Hence compressive sensing exploits this sparse nature of the signal. For simplicity, in this article, we would like to generate the sparse signals which are obtained by utilizing the transformed signal using pseudo-random sequence generator. The uniformly distributed random numbers are selected according to our frame size and using that the K-sparse signal is generated and only the nonzero K values are considered. Different K values are taken for the test and the results are quite similar to any value of K, whether it is less K or greater K.

Considering $x \in R^n$ and the transform coefficient is K-sparse then the measurement m of the basis matrix is selected by generating a random vector uniformly. It is shown in [5, 2] K-sparse vector x can be reconstructed from $y = Ax$ using l_1 minimization provided

$$m \geq CK\ln\frac{n}{K} \tag{11}$$

where $C > 0$ is a universal constant independent of K, n, m. In equation (11), m is directly proportional to K and hence if the sparsity is considered small then the measurement m can also be chosen small in comparison with nso that the solution of an underdetermined system of linear equation is reasonable. Different sparse K signal and the corresponding m measurement by considering $C = 0$ are listed in Table 3.

Recovered signal

The reconstructing can be performed only by the customized speaker which is embedded with the secret key and the security key as shown in Fig. 5(b). The compressed watermarked signal reaches the speaker where the programming recovery takes place using the l_1 minimization with w_k and s_k and the optimum value is obtained by primal dual sparse approximation algorithm. The recovered signal is shown in Fig. 7.

Table 3: Different K and m

S.No	**K**	**m**
1.	3	$\geq$14
2.	5	$\geq$20
3.	7	$\geq$25
4.	10	$\geq$32
5.	13	$\geq$39
6.	15	$\geq$43
7.	20	$\geq$51

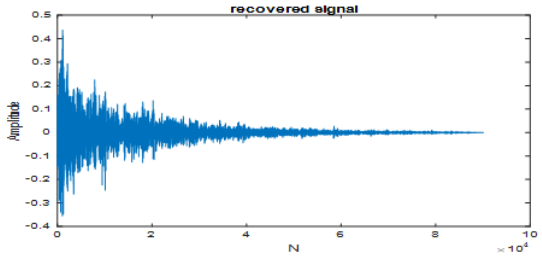

Figure 7: Recovered signal

For our experiment we have tested different instrumental audio data such as piano, guitar etc. and the results are listed below in Table 4. The proposed algorithm takes approximately 2ms to perform a crypto watermarking on an audio of length of 256 samples and takes approximately 0.1s to reconstruct the host signal using the security key and watermark key. Further, the above proposed algorithm is tested on various audio album songs such as Backstreet boys, Emilie Big world and observed that the success rate is around 80%, which yields a good efficiency with a reduced delay for embedding and reconstructing the signal.

9 Imperceptibility

Imperceptibility is the parameter used to measure the perceptual quality of the original audio after embedding the watermark data into it. The objective parameter to measure the imperceptibility is Signal-to-Noise ratio (SNR) and Objective Difference Grade (ODG). The SNR is a measurement that compares the similarity between the undisturbed host signal and the watermarked host signal. The SNR is calculated as

$$SNR = -10 \log_{10} \frac{\sum_{i=1}^{n} (Y - Y')^2}{\sum_{i=1}^{n} (Y)^2} dB \tag{12}$$

Table 4: Test Results

Signal	Sparse K	Measurement m	Successful reconstruction (%)
Piano.wav	3	14	95.2168
		16	95.2684
	5	20	95.1694
		24	95.2339
	10	32	95.0609
		36	95.0795
guitar.wav	3	14	90.6443
		16	90.5757
	5	20	90.6565
		24	90.6322
	10	32	90.3321
		36	90.4074
Equinox.wav	3	14	97.2206
		16	97.1397
	5	20	96.9721
		24	97.0656
	10	32	96.8108
		36	96.8690

where Y is the compressive sensed data without embedding a watermark signature and Y' is the compressive sensed data by embedding the watermark signature.

We have used the kabal [8], PEAQ Basic Model to evaluate the Perceptual Evaluation Audio Quality where $ODG = 0$ means no impairment whereas $ODG = -4$ means it's very annoying. It is observed that the obtained ODG is less than -1.9 which shows the fair perceptual quality of audio.

As the final judgment of the perceptual quality of audio has to be made by the HumanAuditory System (HAS) we have experimented with the subjective quality measurement test also. For the test, we have selected four participants and asked them to grade the dissimilarity between the original host and the recovered signal. The Subjective Difference Grade (SDG) is reported by the participants where $SDG = 5$ means no dissimilar and $SDG = 0$ means totally dissimilar. It is observed that the obtained SDG is greater than four which shows the good perceptual quality of the audio signal. Table 5 shows the SNR, ODG, and SDG of the different audio signals.

Table 5: Imperceptibility measurement

Audio	**SNR**	**ODG**	**SDG**
Piano	32.38	-1.131	> 4
Guitar	32.96	-1.126	> 4
Handel	31.2	-1.889	> 4.5
Bass	31.31	-1.9	> 4.5
440Hz	34.3	-1.32	> 4
1kHz	31.82	-1.2	> 4

10 Robustness

To verify the robustness of the proposed method the following attacks are performed.

a. Amplitude Modification
The amplitude of the watermarked signal is modified by ±6% whereas the positive and negative scale is boosting off the amplitude and cutting off the amplitude respectively.

b. Echo Addition
An echo with a delay of 350ms and echo level of 85% is added to the watermarked audio signal.

c. Filtering
Different filtering such as Low Pass Filter, High Pass Filter, Band Pass Filter and Band Stop Filter with different cut off frequency is applied to the watermarked audio signal.

d. Reverberation
Big room reverberation with a reverberation time of 1000ms is exerted on the watermarked audio signal.

e. Resampling
The watermarked audio is downsampled 22050 Hz and upsampled back to source sampling frequency of 44100Hz.

f. MP3 Compression
The watermarked audio signal is compressed to a bit rate of 16kbps and decompressed back to .wav format.

g. Noise addition
White Gaussian Noise is added to the watermarked audio signal.

To measure the robustness, the commonly used parameters are Normalized Correlation (NC) and Bit Error Rate (BER). The Normalized Correlation (NC) is defined as

$$NC = \frac{x^* \tilde{x}}{\sqrt{x^2}\sqrt{\tilde{x}^2}} \, . \tag{13}$$

The Bit Error Rate (BER) is defined as

$$BER = \frac{x^* \tilde{x}}{n} \tag{14}$$

where x is the recovered signal without any attacks, $\tilde{x}$ is the recovered signal with an attack, and n is the length of the signal. Table 6 shows the NC and BER for the audio files of `Handel.wav` and `guitar.wav`. If $NC = 1$ means the algorithm is high robustness to attacks whereas if $NC = 0$ means the algorithm is fragile to attacks. It can be observed from the table, the proposed algorithm is possessing the nature of high robustness as NC is greater than 0.96 and BER of zero for all cases. The Robustness comparison of the proposed algorithm with the other existing watermarking algorithm is also shown in Table 6.

Table 6: Robustness Test Results of Proposed algorithm and Comparison of Robustness with other watermarking algorithms

	Proposed Algorithm						**Comparison**			
	Handel.wav			Guitar.wav			[8]		[10]	[11]
Attacks	**NC**	**SNR**	**BER**	**NC**	**SNR**	**BER**	**NC**	**BER%**	**BER**	**BER %**
No attack	1	32.37	0	1	36.09	0	1	0	0	0
Amp_6% Boosting	0.972	32.14	0	0.979	35.55	0	*NR	NR	0	0
Amp_6% Cut	0.969	31.46	0	0.977	34.28	0	NR	NR	0	0
Echo Addition	0.969	32.42	0	0.985	35.55	0	NR	NR	0.004	0.01
Filter_LPF	0.978	31.99	0	0.969	33.12	0	0.948	6	0	0
Filter_HPF	0.963	31.83	0	0.974	35.56	0	NR	NR	NR	0
Filter_BPF	0.979	32.57	0	0.969	33.65	0	Not Reported			
Filter_BSF	0.961	32.09	0	0.975	37.08	0	Not Reported			
Reverberation	0.968	33.96	0	0.972	34.56	0	Not Reported			
Resampling	0.960	31.53	0	0.961	33.50	0	0.978	3	0	0
MP3Compression	1	33.56	0	1	36.09	0	0.987	1	0	0
Noise	0.999	32.37	0	0.999	35.84	0	1	NR	0.15	0.01

*NR – Not Reported

11 Comparison

The proposed algorithm in this article is compared with the recent audio watermarking scheme. Each algorithm uses different properties and we have chosen SNR, ODG and SDG values as the comparisonparameter with our proposed algorithm. All the compared algorithms, embed the watermark in the digital data and reported SNR values is greater than 20 dB whereas the reported ODG is less than -2. Comparing with the other methods, our method proposes a high SNR which is greater than 31dB. As we use the crypto watermarking at the time of ADC, the ODG values observed is fair compared with the other method. We can make a convenient tradeoff in this case as the watermark is embedded at the time of signal acquisition. Table 7 shows the comparison of a different watermarking algorithm.

Table 7: Comparison with other Watermarking Algorithm

Algorithm	**SNR (dB)**	**ODG**	**SDG**
Guo et al. (2012)	20	Not reported	Not reported
Cairong Li et al. (2012)	22.35 to 27.35	Not reported	Not reported
Khaldi and Boudraa (2013)	24.12 to 26.38	0.4 to -0.6	Not reported
Arnold et al. (2014)	0 to 25	-0.42	-1.07
Hu et al. (2014)	20.889	-0.062	Not reported
Fallahpour, Megias (2015)	35 to 61	-0.3 to -1.1	> 3.5
Proposed Algorithm	31.2 to 34.3	-1.1 to -1.9	> 4

12 Conclusion

The proposed crypto-watermarking algorithm is based on compressive sensing and by exploiting a partially decomposed Haar matrix, the watermark is generated. The results show the SNR is above 30dB which shows that the perceptual quality of the audio is not degraded in the name of increasing the security. The security of the audio is more as the watermark and security key are embedded into the host audio signal at the time of signal acquisition only. Hence the proposed algorithm can be utilized for real-time application and can be used to protect the original audio from illegal copying. The results of the robustness shows that the NC is close to unity and BER is zero and therefore the algorithm is highly robust against various signal attacks such as noise addition, echo addition, reverberation, etc. Hence the proposed algorithm can be used to embed a watermark in a live concert and protect the data by providing the security key.

References

[1] Arnold, Michael, Chen, Xiao-Ming, Baum, Peter, Gries, Ulrich, and Doerr, Gwenael. A phase-based audio watermarking system robust to acoustic path propagation. *IEEE Transactions on Information Forensics and Security*, 9(3):411–425, 2014. DOI: 10.1109/TIFS.2013.2293952.

[2] Candes, E.J. and Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008. DOI: 10.1109/MSP.2007.914731.

[3] Candes, Emmanuel, Romberg, Justin, and Tao, Terence. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. DOI: 10.1002/cpa.20124.

[4] Fallahpour, Mehdi and Megias, David. Audio watermarking based on Fibonacci numbers. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(8):1273–1282, 2015. DOI: 10.1109/TASLP.2015.2430818.

[5] Foucart, Simon and Rauhut, Holger. *A Mathematical Introduction to Compressive Sensing*. Springer Science+Business Media, Birkhäuser, New York, NY, 2013. DOI: 10.1007/978-0-8176-4948-7.

[6] Guo, Qijun, Zhao, Yanbin, Cheng, Pingpan, and Wang, Fengming. An audio digital watermarking algorithm against A/D and D/A conversions based on DCT domain. In *Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pages 871–876, Yichang, China, 2012. IEEE. DOI: 10.1109/CECNet.2012.6201522.

[7] Hu, Hwai-Tsu, Chou, Hsien-Hsin, Yu, Chu, and Hsu, Ling-Yuan. Incorporation of perceptually adaptive QIM with singular value decomposition for blind audio watermarking. *EURASIP Journal on Advances in Signal Processing*, 2014. DOI: 10.1186/1687-6180-2014-12.

[8] Kabal, P. An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality. Technical report, McGill University, 2002.

[9] Khaldi, Kais and Boudraa, Abdel-Ouahab. Audio watermarking via EMD. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(3):675–680, 2013. DOI: 10.1109/TASL.2012.2227733.

[10] Li, Cairong, Hu, Ruimin, and Zeng, Wei. Radon transform and DWT based audio watermarking algorithm against DA/AD conversion. In *Proceedings of the International Conference on Audio, Language and Image Processing*, pages 282–286, Shanghai, China, 2012. IEEE. DOI: 10.1109/ICALIP.2012.6376626.

[11] Lin, Yiqing and Abdulla, Waleed H. *Audio Watermark: A Comprehensive Foundation Using MATLAB*. Springer International Publishing, Switzerland, 2015. DOI: 10.1007/978-3-319-07974-5.

[12] Mishali, M., Eldar, Y.C., Dounaevsky, O., and Shoshan, E. Xampling: Analog to digital at sub-Nyquist rates. *IET Circuits, Devices & Systems*, 5(1):8–20, 2010. DOI: 10.1049/iet-cds.2010.0147.

[13] Qi, Jin, Hu, Xiaoxuan, Ma, Yun, and Sun, Yanfei. A hybrid security and compressive sensing-based sensor data gathering scheme. *IEEE Access*, 3:718–724, 2015. DOI: 10.1109/ACCESS.2015.2439034.

[14] Selesnick, Ivan. Introduction to sparsity in signal processing, 2012. Connexions Web site, http://cnx.org/content/m43545/.

Received 5th March 2019