Indonesian Journal of Electrical Engineering and Computer Science
Vol. 18, No. 2, May 2020, pp. 1096~1103
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i2.pp1096-1103

Securing speech signals by watermarking binary images in the wavelet domain

Rakan Saadallah Rashid[1], Jafar Ramadhan Mohammed[2]
[1]Bardarash Technical Institute, Duhok Polytechnic University, Kurdistan Region of Iraq
[2]College of Electronic Eng, Ninevah University, Iraq

Article Info

Article history:

Received Aug 6, 2019
Revised Sep 23, 2019
Accepted Oct 23, 2019

Keywords:

Audio Signal
Binary Image
Watermarking
Wavelet Transform

ABSTRACT

Digital watermarking is the process of embedding particular information into other signal data in such a way that the quality of the original data is maintained and secured. Watermarking can be performed on images, videos, texts, or audio to protect them from copyright violation. Among all of these types of watermarking, audio watermarking techniques are gaining more interest and becoming more challenging because the quality of such signals is highly affected by the watermarked code. This paper introduces some efficient approaches that have capability to maintain the signals' quality and preserves the important features of the audio signals. Moreover, the proposed digital audio watermarking approaches are performed in the transform domain. These approaches are gaining more attention due to their robustness or resistance to the attackers. These transform domains include discrete cosine transform (DCT), short-term Fourier transform (STFT), and digital wavelet transform (DWT). Furthermore, the most digital wavelet transforms were found to be applicable for speech watermarking are the Haar and the Daubechies-4.

Corresponding Author:

Rakan Saadallah Rashid,
Bardarash Technical Institute,
Duhok Polytechnic University, Duhok, Kurdistan Region of IRAQ
Email: ijeecs.iaes@gmail.com

1. INTRODUCTION

Nowadays, the amount of digital information is becoming huge and it can be easily distributed everywhere in the world by means of the Internet or computer networking. This digital information could be in any of the following forms: text, video, images, and audio. The audio signals such as speech and music signals are considered in this paper. These signals can be copied, duplicated, and/or digitally manipulated. To keep up with the transmission of digital information over the Internet, the reliability and originality of the transmitted information should be maintained. Thus, it is necessary that the available information should be protected and secured, especially in some crucial applications that need to be highly secured such as biomedical signals and speech recognition systems. One method of coping with this problem is by embedding an invisible watermark code into the original information to mark the ownership of it. Generally, there are many methods of information protection, which can be divided into different categories such as cryptography, steganography, and watermarking [1-3].

There are several techniques in which a watermark code is embedded in an original audio file. The most widely used technique involves the watermark process in either the time domain (also called the special domain in some references) or the frequency domain (or transform domain). Generally, the performances of the transform domain approaches are much better than those of the special domain approaches [4-5]. The commonly used transform domain approaches are based on discrete cosine transform (DCT) [6], short-term Fourier transform (STFT) [7], discrete wavelet transform (DWT) [8] and cepstrum transform [8].

The watermarking approaches based on the advanced transform domain, specifically the wavelet domain of the audio signal, are adopted in this research work. WT is one of the most popular frequency domain examples [9]. In the literature, the watermarking approaches that depend on the spatial domain were introduced first. They are simple and easy to design. For example, Cai and Gopalan [10], proposed an approach that depends on the spatial domain. This watermarking approach depends on bit modification of voiced or unvoiced segments, where five consecutive samples in each voiced or unvoiced segment of the original audio signal are selected and replaced by the watermark code. The authors have proved the effectiveness of this approach and the attacker cannot view the watermark code. However, the robustness was found to be insufficient in some applications. Therefore, many researchers have concentrated their studies on the approaches based on the transform domain rather than simple space domains. The methods that were characterized as transform domains are classified as high-robustness watermark approaches. As a result, transform domain approaches were developed. First, Khayam [11], 2003, used DCT to transform a signal in the spatial domain into the frequency domain.

Then, Chen, Wen-Yuan; Huang., Shih-Yuan [12], suggested a watermarking approach that uses DCT to embed a watermark code into original image data. The authors investigated the performance of the described approach and compared it to its spatial domain counterpart. The principle role of DCT is to divide the original data signal into three band frequencies (i.e. lower-frequency band, middle-frequency band, and high-frequency band). The authors use the lower-frequency band to embed the watermark code.

Recently, some other researchers were used DWT to perform watermarking on various applications [13-16]. However, most of the used methods were found applicable to the images rather than speeches. In this paper, the watermarking of the speech signals in the wavelet transforms is performed. Moreover, some performance measures are introduced to study the effect of watermarking on the signal’s quality.

Nin and Ricciardi [17], concentrated their study on digital watermarking and security-related issues in the information society. A study of the effect of MPEG compression on audio watermarking was carried out by Artameeyanant [18], who explored the principle of human hearing called the psychoacoustic model.

Elshazly, Fouad, and Nasr [19], tried to enhance the security and robustness of audio watermarking. The described approach is based on mean-quantization in DWT. Namazi, Karami, and Ramazannia [20], in this approach, the watermark code was embedded into the middle-frequency band of the original data signal. The middle-frequency band was chosen so that the robustness would be guaranteed when its performance against some signal processing operations such as compression was investigated. The original data signal was partitioned into many blocks and then DCT was performed on each block. The middle frequency band was selected and the watermark code was embedded in the assigned frequencies.. Kamaladas and Dialin [21], this method extracts the acoustic features of the speech signal and saves it as a fingerprint in a database. By using the fingerprint-matching techniques, the speech signals can be successfully identified. Lai [22], used SVD in watermarking. This approach uses some visual features of the original data to find out the proper location of the watermark code and shows a good performance. Singh, Dave, and Mohan [23], propose an approach to decompose the original information using wavelet transform. The authors showed that this approach can provide good robustness against some signal-processing operations and maintain a high quality of the watermarked signal at the same time. Both the watermark information and the hidden key have to be selected and must be available prior to the embedding stage [24-25].

One of the most important practical implementations of the digital watermarking is copyright protection. The second application is the owner identification. As a proof of ownership, some important information regarding the identification of the owner may be inserted as a watermark code into the original signal. The applications of the owner identification need a high level of security. The third application is copy control. This application has a unique feature to prevent unwanted copying processes through a consumer-control mechanism. This control mechanism prevents unwanted or illegal copying and recording of the information by inserting a special watermark code called a “never-copy” watermark. It can also limit the number of times the information can be copied. The fourth application is broadcast monitoring. The media and the advertisers use this type of application to be sure that broadcasters will air commercials at the time and location that they want according to the contracts. Watermark codes can be inserting into any form of digital information for broadcasting on network systems.

Another application is in biomedicine. In these applications, the watermarking code can be the full name of the patient and some other related personal information, which are unique on X-ray reports or MRI scan reports. This application is very important because it avoids or reduces the misplacement of medical reports, which are very important during treatment.

There are also some applications in airports or airline monitoring. The use of watermarking approaches may also be effective in applications of airline-traffic monitoring, where the pilot may communicate with a ground station system via a voice or speech signal. This communication can easily attack. This may result in a breakdown of the communication between the pilot and the ground station. In order to overcome such problems, the flight number may be embedded into the speech signal during the communication between the ground station and the flight pilot. It is well known that the flight numbers are unique, and thus the tracking of flights will become more secure and easy.

2. THE PROPOSED APPROACHE

In this work, we mainly focus on the process of hiding a watermark code such as a binary image in an audio signal such as a speech or music signal. Here, only binary images are considered as watermark codes, while all other types of images such as JPEG and MPEG are not considered. This is mainly because binary images contain a smaller number of pixels (where the values of the pixels are either 0 or 1 and the total number of the pixels may be easily chosen during the creation process of the binary image). By hiding these pixels in the recorded speech signal in a proper way, the original recorded speech signal will not much affected and its quality will remain at an acceptable level. That is why we only considered binary images. Moreover, in this work, the process of creation of specific binary images is described. Also, the process of hiding those binary images in the wavelet domain of the recorded speech signals is also presented. The described watermarking approach consists of the following steps:

Step 1: The required audio signal that needs to be watermarked for the purpose of protection is prepared. This audio signal is recorded by Matlab software and sampled with a specific sampling rate equal to 11025 Hz. In order to achieve this step, the following Matlab commands are used.

```
clear all; clc; close all;
Fs=11025;% sampling rate
y = wavrecord(n,Fs)
wavwrite(y,Fs,'D:\Cankaya.wav')
wavplay(y,Fs)
```

On the other hand, the binary image is created by the following Matlab command

```
I=imread('rakan.bmp');
imshow(I)
title('Binary Image');
```

Step 2: Then the recorded audio signal is divided into segments, where each segment consists of 4487 samples. Every segment of the recorded speech signal is transformed or decomposed using either Haar or Daubechies-4 wavelet transformation or any other transformation. The process starts by inserting the first pixel of the binary data collected from the binary image into the third-level detail coefficients that are obtained using wavelet decomposition of the first segment of the recorded speech signal. This is done by changing the values of the third-level detail coefficients such that the mean or average of the third-level detail coefficients is modified to a new value. This new value may be represented by m+e if the binary value is ‘1’ or a value of m – e if the binary value is ‘0’. Here the parameter m represents the mean of the third-level detail coefficients and the parameter e represents a constant number. Its value is chosen to be equal to the value of one fifth of the energy of the third-level detail coefficients. Note that the value of the variance should remain the same so that the quality of the recorded speech signal is preserved.

Step 3: This process is repeated for the other segments of the recorded speech signal in order to hide the complete pixels obtained from the binary image.

Step 4: After completing the above steps, the result is that the pixels of the required watermark code or binary image will be completely hidden in the wavelet domain of the recorded speech signal.

Step 5: For the purpose of identification, the hidden watermark binary code can be recovered easily by comparing the mean of the corresponding detail coefficients computed before and after hiding. If the value of the average of the third-level detail coefficient of the specific segment corresponding to the original recorded speech signal is greater than the average of the third-level detail coefficients of the respective segment corresponding to the watermarked speech signal, the value of the pixel will be considered as ‘1’; otherwise, its value will be considered as ‘0’. All these estimated values of the pixels will be stored to construct or recover the hidden watermark code.

Note that the total number of samples (or period) of the original recorded speech signal is chosen such that all the binary pixels collected from the binary image are hidden properly in the recorded speech signal. Figure 1 shows the flow chart of the Matlab implementation.

For comparison, other transforms such as DCT [11] and STFT [6] were also implemented. To extract the pixels of the binary image from the watermarked audio signal, first the watermarked and the original recorded audio signals are transformed using the wavelet transform. Second, the transformed original audio signal is subtracted from the transformed watermarked audio signal. This is because the binary image is the difference between the original audio signal and the watermarked audio signal. Finally, the similarity between the original watermark code (binary image) and the extracted watermark code is evaluated.

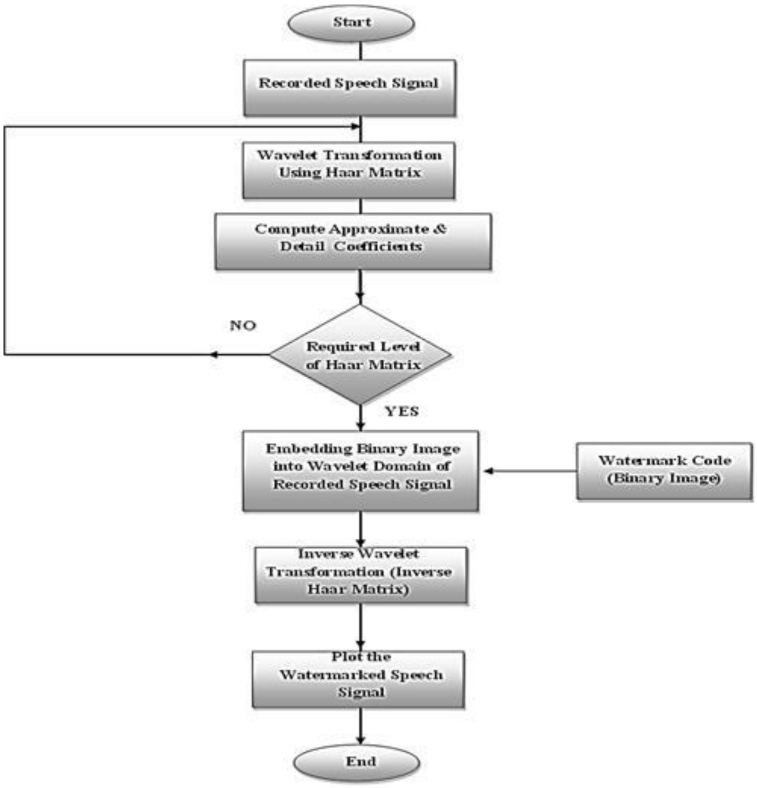

Start
Recorded Speech Signal
Wavelet Transformation Using Haar Matrix
Compute Approximate & Detail Coefficients
NO
Required Level of Haar Matrix
YES
Embedding Binary Image into Wavelet Domain of Recorded Speech Signal
Watermark Code (Binary Image)
Inverse Wavelet Transformation (Inverse Haar Matrix)
Plot the Watermarked Speech Signal
End

Figure 1. The flowchart of the Matlab program used to implement the proposed watermark approach

3. RESULTS AND ANALYSIS

In order to validate the performance of the described approach of watermarking, extensive computer simulations carried out by means of Matlab software are presented in this section. Regarding the audio signals, this research work used recorded speech signals of 2 s length. Moreover, some music signals were also used. Generally, the sampling frequency was chosen to be equal to 11025 Hz. The original recorded speech and music signals were decomposed up to the fourth level using either Haar or Daubechies-4 wavelet matrix transformations and both approximate and detail coefficients were selected to embed the pixels of the watermark code or binary image. Here, the first binary image is like a square geometrical shape. Figure 2 shows the original recorded speech signal of a sentence or word "rakan". This figure also shows the watermarked speech signal after embedding the binary image or watermark code. The difference between the original and watermarked speech signals is also displayed in this figure. For this particular example, the watermark code or binary image of 45 × 45 pixels is shown in Figure 3.

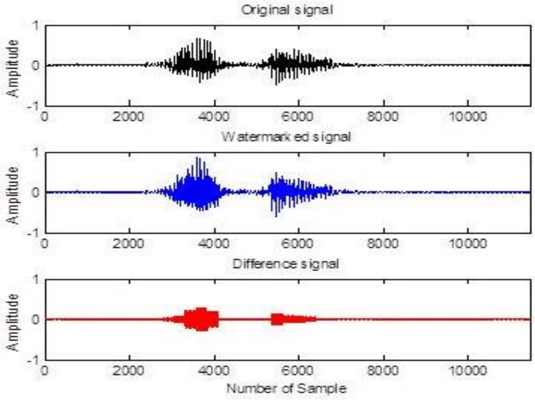


Figure 2. Speech signal "rakan" before and after Watermarking with the binary image "square"

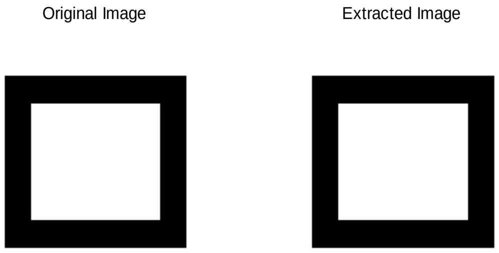


Figure 3. Getting the binary image from the watermarked speech signal and compare it with original one

Figure 4 shows the results of watermarking a speech signal of a sentence or word "cankaya" with the binary image "square". In another experiment, we record a music signal and we want to watermark it. The music signal that is used is a melody signal, which is an audio signal but is completely different from a speech signal. This melody signal is played and we used Matlab commands to record it. In this case, the binary image is chosen to be in the shape of "rakan". Figures 5 and 6 show the results of watermarking a melody music signal with the binary image "rakan".

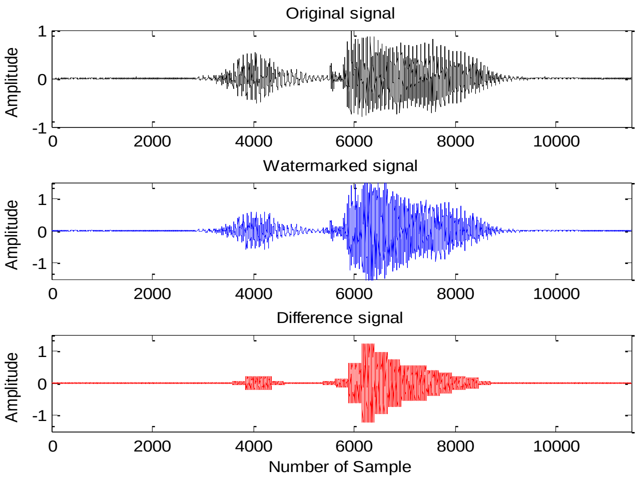


Figure 4. Speech signal "cankaya" before and after watermarking with the binary image "square"

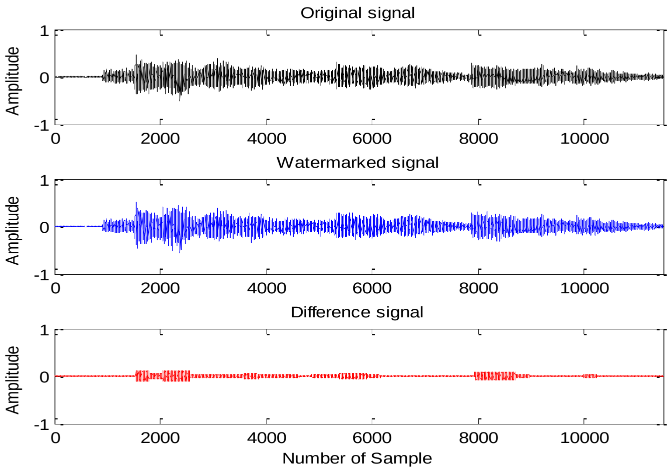


Figure 5. Music signal "melody" before and after watermarking with the binary image "rakan"

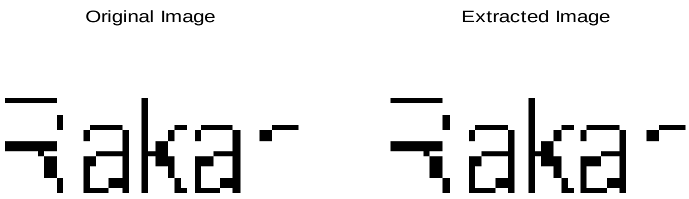


Figure 6. Getting the binary image from the watermarked music signal and compare it with original one

The signal shown in the top of Figure 5 represents the recorded melody signal without the watermark code, while the middle signal represents the watermarked signal (the output signal of the proposed approach). The signal on the bottom of Figure 5 represents the difference between the recorded speech signal and the watermarked speech signal. This difference signal is one way to show the performance of the proposed watermark approach. For good performance, its amplitude should be as small as possible but it can never be zero since the pixels of the binary image have been added to the samples of the recorded speech signal.

The figure on the left side of Figure 6 represents the binary image that will be used for the purpose of watermarking in the embedding process, while the figure on the right side represents the watermark code extracted during the extraction process. It can be seen that the two figures (on the left and right sides) are identical to each other. This is exactly true for the case where none of the signal processing operation attackers were considered. On the other hand, the extracted watermark code will be slightly different from the original one when considering some attackers such as AWGN.

In order to fully validate the performance of the proposed approach, it should be compared with some other existing watermarking approaches. In addition, it is also necessary to investigate the performance of the proposed watermarking approach under different signal processing operations (attackers). The most common attackers considered in this paper are AWGN, filtering, and compression.

For comparison purposes, we need to define some essential performance measures. The most commonly used measures are the Signal-to-Noise Ratio (SNR) and Normalized Correlation (NC).The SNR evaluates the quality of the watermarked audio signal, while the NC measures the similarity (or correlation) between the original watermark code (the created binary image) and the extracted watermark code. In mathematical equations, these two performance measures are defined as [9]:

$$\mathrm{SNR} = 10\log\frac{\sum_{n=0}^{N-1} S(n)^2}{\sum_{n=0}^{N-1}[S(n)-S(\acute{n})]^2} \tag{1}$$

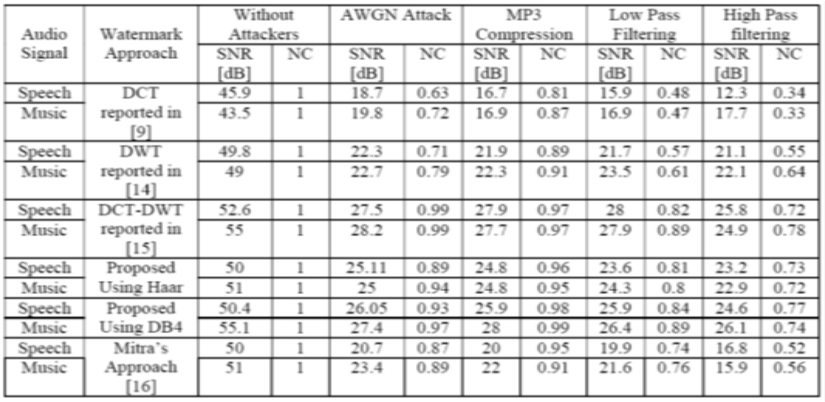
Where $S(n)$ represents the original recorded speech or music signal, $S(\acute{n})$ represents the watermarked speech or music signal (i.e., the output of the proposed watermark approach), and $S(n) - S(\acute{n})$ represents the difference signal that was plotted previously in Figures 3, and 5. The variable N stands for the total number of samples in the recorded speech signal, which is chosen to be 11448 by Matlab.

$$\mathrm{NC} = \left|\frac{\sum_{i=1}^{I}\sum_{j=1}^{J} B(i,j)*B(\acute{\imath},\acute{\jmath})}{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(B(i,j)*B(\acute{\imath},\acute{\jmath})\right)^{0.5}}\right| \tag{2}$$

Where $B(i,j)$ represents the original watermark code created (binary image), $B(\acute{\imath},\acute{\jmath})$ represents the hidden watermark code which is recovered during the extraction process, and the variables I and J stand for the pixel's location. It is worth mentioning that, in the ideal case (no attackers), the value of the NC is 1 and the value of the SNR is about 50 dB. Table 1 shows the results of these performance measures for various watermarking approaches.

From this table, it can be seen that the best values for SNR and NC were obtained under the case of no attach or operation. These values were noticed to significantly reduce when considering any of the signal processing operations (AWGN, MP3, Low pass and High pass filtering). Also note that for AWGN case, the results of the proposed approach with Haar and dB4 are outperformance with compared to all other approaches. It is also found that the performance of the DWT is better than the DCT [9] and Mitra’s [14] algorithms.

Table 1. Performance Comparison of Different Approaches under Various Signal Processing Operations (Attackers)



Audio Signal	Watermark Approach	Without Attackers		AWGN Attack		MP3 Compression		Low Pass Filtering		High Pass filtering	
		SNR [dB]	NC	SNR [dB]	NC	SNR [dB]	NC	SNR [dB]	NC	SNR [dB]	NC
Speech	DCT reported in [9]	45.9	1	18.7	0.63	16.7	0.81	15.9	0.48	12.3	0.34
Music		43.5	1	19.8	0.72	16.9	0.87	16.9	0.47	17.7	0.33
Speech	DWT reported in [14]	49.8	1	22.3	0.71	21.9	0.89	21.7	0.57	21.1	0.55
Music		49	1	22.7	0.79	22.3	0.91	23.5	0.61	22.1	0.64
Speech	DCT-DWT reported in [15]	52.6	1	27.5	0.99	27.9	0.97	28	0.82	25.8	0.72
Music		55	1	28.2	0.99	27.7	0.97	27.9	0.89	24.9	0.78
Speech	Proposed Using Haar	50	1	25.11	0.89	24.8	0.96	23.6	0.81	23.2	0.73
Music		51	1	25	0.94	24.8	0.95	24.3	0.8	22.9	0.72
Speech	Proposed Using DB4	50.4	1	26.05	0.93	25.9	0.98	25.9	0.84	24.6	0.77
Music		55.1	1	27.4	0.97	28	0.99	26.4	0.89	26.1	0.74
Speech	Mitra's Approach [16]	50	1	20.7	0.87	20	0.95	19.9	0.74	16.8	0.52
Music		51	1	23.4	0.89	22	0.91	21.6	0.76	15.9	0.56

4. CONCLUSION

Audio watermarking is a newly advancing field of study and its importance is rapidly increasing. Watermarking can be applied to a various application including speech signals, music, and biomedical signals. It can be performed in either time or transform domains. Some common transforms include DCT, STFT, and DWT. The digital wavelet transforms include Haar and Daubechies-4. The presented results show that the DWT is capable to preserve the quality of the watermarked audio signals better than other approaches. In addition, the watermark code (binary image) was recovered during the extraction process with an acceptable value of the normalized correlation (NC) even under a sever condition like additive white Gaussian noise (AWGN), low pass and high pass filtering, and MP3 compression. In view of the above, the proposed watermarking approach was found to be robust, secure, and effective.

REFERENCES

[1] Li X., Zhang M., & Zhang R., "*A New Adaptive Audio Watermarking Algorithm*", in Proceedings. 5th World Congress on Intelligent Control and Automation, WCICA 4, Hangzhou, China, pp. 4357-4361. 2004.

[2] Abbasfard M., "Digital Image Watermarking Robustness: A Comparative Study", M.Sc. thesis, Department of Computer Engineering, Delft University of Technology, The Netherlands. 2009.

[3] Zebari, D. A., Haron, H., Zeebaree, S. R., & Zeebaree, D. Q. (2018, October). "*Multi-Level of DNA Encryption Technique Based on DNA Arithmetic and Biological Operations*". In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 312-317). IEEE.

[4] Sruthi N., Sheetal A. V., &Elamaran V., "*Spatial and Spectral Digital Watermarking with Robustness Evaluation*", in Proceedings of the International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), pp. 500-505. 2014.

[5] Brifcani, A. M. A., & Brifcani, W. M. A. (2010). "Stego-based-crypto technique for high security applications". *International Journal of Computer Theory and Engineering*, 2(6), p. 835.

[6] Jiansheng M., Sukang L., &Xiaomei T., "*A Digital Watermarking Algorithm Based on DCT and DWT*", in IEEE Proceedings of the International Symposium on Web Information Systems and Applications, Nanchang, P. R. China. pp. 67-75. 2009.

[7] Sonoda K., &Sek A., "*Digital Watermarking Method Based on STFT Histogram*", in IEEE Proceedings of the Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 287-290. 2013.

[8] Tang X., Niu Y., Yue H., & Yin Z., "*A Digital Audio Watermark Embedding Algorithm with WT and DCT*", in Proceedings of MAPE 5, Beijing, China, pp. 970-973. R2 7. Kaengin S., Pathoumvanh S., &Airphaiboon S., (2008), "*Binary Image Watermarking on Audio Signal Using Mean-Quantization and Wavelet Transform*", in Proceedings of ISBME 2008, Bangkok, Thailand, pp. A-18. 2005.

[9] Zebari, D. A., Haron, H., Zeebaree, S. R., & Zeebaree, D. Q. (2019, April). "*Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform*". In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 100-105). IEEE.

[10] Cai D., &Gopalan K., "*Audio Watermarking Using Bit Modification of Voiced or Unvoiced Segments*", in IEEE Proceedings, pp. 501-506. 2014.

[11] Khayam S. A., "The Discrete Cosine Transform (DCT): Theory and Application", Michigan State University. pp. 203-209. 2003.

[12] Chen W-Y & Huang S.-Y., "Digital Watermarking Using DCT Transformation", pp.98. Heidelberg, pp. 252-266. 2000.

[13] Wenhuan Lu, Zonglei Chen, Ling Li, Xiaochun Cao, Jianguo Wei, Naixue Xiong, Jian Li, and Jianwu Dang, "Watermarking Based on Compressive Sensing for Digital Speech Detection and Recovery," *Sensors 2018*, 18(7), 2390; https://doi.org/10.3390/s18072390.

[14] Wen Diao,Yuanxin Wu, Weiming Zhang, Bin Liu, and Nenghai Yu, "*Robust Audio Watermarking Algorithm Based on Air Channel Characteristics*," 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC) (2018), Guangzhou, China, Jun 18, 2018 to Jun 21, 2018, ISBN: 978-1-5386-4210-8, pp: 288-293.

[15] Brifcani, A. M. A., & Al-Bamerny, J. N. (2010, December). "*Image compression analysis using multistage vector quantization based on discrete wavelet transform*". In 2010 International Conference on Methods and Models in Computer Science (ICM2CS-2010) (pp. 46-53). IEEE.

[16] KS, A., & Dharun, V. S. "A Pepy Algorithm For Blind Watermarking Of Digital Images Based On Dwt And Cryptography".

[17] Nin J., & Ricciardi S., (2013), "*Digital Watermarking Techniques and Security Issues in the Information and Communication Society*", in IEEE Proceedings of the 27th International Conference on Advanced Information Networking and Applications Workshops, pp. 1553-1558.

[18] Artameeyanant P., (2010), "Wavelet Audio Watermark Robust Against MPEG Compression", SICE Annual Conference, Kagawa University, Japan, pp. 1375– 1378.

[19] Elshazly A. R., Fouad M. M., & Nasr M. E., (2012), "*Secure and Robust High Quality DWT Domain Audio Watermarking Algorithm with Binary Image*", in IEEE Proceedings, pp. 207-212.

[20] Namazi F., Karami M. R., & Ramazannia S. B., (2012), "Block Based Adaptive Image Watermarking Scheme Using Visual Perception Model in DCT Domain", *International Journal of Computer Applications*, vol. 4, pp. 41-45.

[21] Kamaladas M. D., & Dialin M. M., (2013), "*Fingerprint Extraction of Audio Signal Using Wavelet Transform*", in IEEE Proceedings of the International Conference on Signal Processing, Image Processing and Pattern Recognition (ICSIPR), pp. 308-312.

[22] Lai C. C., (2011), "An Improved SVD-Based Watermarking Scheme Using Human Visual Characteristics", *Optics Communications*, vol. 14, pp. 938-944.

[23] Singh A. K., Dave M., & Mohan A., (2013), "*Performance Comparison of Wavelet Filters against Signal Processing Attacks*", in IEEE Proceedings of the 2nd International Conference on Image Information Processing, pp. 695-698.

[24] Aniruddha K. and Aghila G., "Robust image-in-audio watermarking technique based on DCT-SVD transform," *EURASIP Journal on Audio, Speech, and Music Processing*, October 2018.

[25] Zebari, D. A., Haron, H., Zeebaree, S. R., & Zeebaree "Security Issues in DNA Based on Data Hiding: A Review". *International Journal of Applied Engineering Research*, 12(24), pp. 15363-15377, 2017.