IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Audio Watermarking for Security and Non-Security Applications

(December 2021)

Maha Charfeddine[1], Member, IEEE, Eya Mezghani[1], Salma Masmoudi[1] , Chokri Ben Amar[2] and Hesham Alhumyani[2]

[1] REGIM: REsearch Group on Intelligent Machines, University of Sfax, National School of Engineers (ENIS), BP 1173, Sfax, 3038, Tunisia
[2] Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Corresponding author: Maha Charfeddine (maha.charfeddine.tn@ieee.org)

This research was supported by Taif University Researchers Supporting Project Number (TURSP-2020/216), Taif University, Taif, Saudia Arabia

ABSTRACT The digitization of audiovisual data is significantly increasing. Thus, in order to guarantee principally the protection of intellectual properties of this digital content, watermarking has appeared as a solution. The watermarking can be used in reality in several types of applications that target two different contexts; the first for security applications and the second for non-security ones. In this paper, we carry a big interest in studying these two types of applications. Moreover, we propose a first digital watermarking scheme for security copyright protection application where we have involved Neural Network architecture in the insertion and detection processes and we have integrated some masking phenomena of the Human Psychoacoustic Model with Linear Predictive Coding spectral envelope estimation of the audio file. Experimentations proved the efficiency of exploiting perceptual masking with spectral envelope consideration in terms of imperceptibility and robustness results. Besides, we suggest a second audio watermarking technique for non-security content characterization application based on deep learning classification architecture. In this scheme, extracted watermark will advise about the audio class: music or speech, the speaker gender and emotion. Reported results indicated that the suggested scheme achieved higher performance at classification level as well as at watermarking properties.

INDEX TERMS Copyright Protection, Human Psychoacoustic Model, Linear Predictive Coding, Audio Content Characterization, Deep Learning Architecture.

I. INTRODUCTION

Information, by way of an expression of knowledge, is seemingly the most valuable asset of humanity. The digitalization advent, delivered us a number of easy-to-use and reasonably cost-free channels to transfer ideas and exchange information. Nonetheless, the instantaneous effect of digitization has been a proliferation of illegal copying that involve violating intellectual property rights. To resolve these problems, a digital watermark can be hidden in a piece of digital content that may comprise audit-trail or copy-limitation information to help copyright enforcement [1]. Digital watermarking offers great chances for not only protection of copyrighted data, but also serves as a general framework to embed information within generic data sorts for various usages. In this paper, we explain several digital watermarking usages and we classify them into security and non-security applications. Next, we introduce two digital watermarking techniques, which can consider basic (standard content) or sensitive data (political news, Quranic data, audio records, confidential communication, etc.). These two watermarking techniques operate distinctively in security and non-security contexts.

This paper is planned as follows: section two presents a definition of digital watermarking then explores its security and non-security applications with some previous works we have already developed in such fields. Section 3 exhibits two proposed audio watermarking schemes in both security and non-security usages for standard and sensitive audio contents. Finally, conclusion is presented in the last section with perspectives for future researches.

II. WATERMARKING DEFINITION AND APPLICATIONS

Watermarking system principally involves two parts; embedding and extraction processes. They use generally a cryptographic key, that could be a public key or a secret key. Watermark is the signature hidden on the original digital content. Watermarked document is an output data resulted by superimposing the original document and the signature. Watermark embedding is displayed in Figure 1 while extraction process is shown in Figure 2.

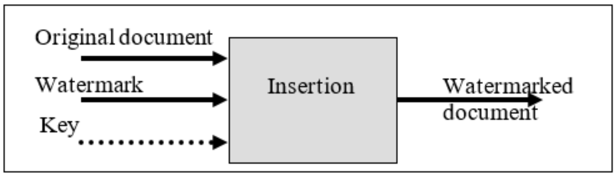


FIGURE 1. **General digital watermark embedding process**

Watermark, original digital content and sometimes the key are set as the input to the embedding process. One basic requirement to differentiate between watermarking techniques is the insertion domain [2-5]: insertion domain with no transformation, the frequency domain and the multi-resolution one.

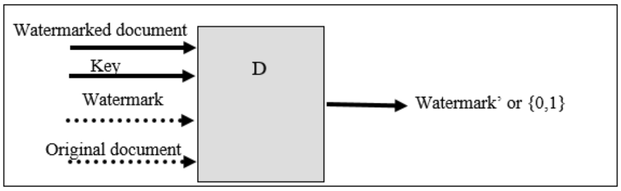


FIGURE 2. **General digital watermark extraction process**

If the original document is not required for the detection step, then the watermarking scheme is blind [6] else, it is non-blind [7]. The performance of watermarking systems entails a number of properties, some of them are:

- Imperceptibility: it is the most important criterion for a digital watermarking. However, we can retrieve in the literature some watermarking techniques that hide perceptible watermark [8-10].
- Robustness: it means that the hidden watermark in a data can endure different attacks and modifications. In most circumstances, we would like that the watermark is robust [2, 11, 12]. In other cases, we wish that any processing in the watermarked document jumps the signature [13-15]. Finally, we wish an intermediate situation, where the mark persists in spite of some processing, and not for others: we call it semi-fragile watermarking [16, 17].
- Security: Only the legal users can extract the watermark and therefore a proprietor can reach the goal of copyright protection.
- Capacity: it defines maximum amount of data that can be hidden into a digital document. This capacity is habitually significant, as many systems need a great payload to be hidden.

Digital watermarking has more than one application. We choose in this paper the following classification of watermarking applications:

A. SECURITY WATERMARKING APPLICATIONS

In the security usage, watermarking aims to adjust the hidden mark according to the action on the digital content held by the hacker. In this case, embedded information must be robust to different intentional piracy attacks. Among the security watermarking applications, we notice:

1) DATA HIDING

The well-known application where data is embedded and transmitted secretly in such a way that no illicit person can discover it [18, 19].

2) SURREPTITIOUS COMMUNICATIONS

Principally steganography applications in military, where persons would like to send secret messages to each other without being perceived [20, 21].

3) OWNERSHIP PROOF

To avoid the unlawful alteration of digital data, the lawful individual credentials is embedded into the digital content [22, 23].

4) AUTHENTICATION

The data can be simply interfered without nay being detected. The signature can be hidden to avoid this tamper and to preserve consequently its originality. For example, the interference of a digital image can easily be discerned because the pixel value of the inserted data would modify and not conform the original one. [24, 25].

5) PROPRIETOR IDENTIFICATION

It is somewhere written on the wrapper of an object such as, identification brand of the paper maker. These kinds of watermark can be effortlessly removed by cropping or tearing the paper. Thus, to overcome this problem, watermark bits identifying the owner, are hidden forming then an integral part of the digital content [26, 27].

6) COPYRIGHT PROTECTION

The proprietor can hide the signature in the data for the protection of conspicuous content. There always has been a problem in supplying the owner identity of an object. In addition, if there is a disagreement concerning the data proprietorship then the owner identity can be effortlessly extracted from the watermark [28, 29].

In this context, we have already developed in [30], an audio watermarking scheme where the watermark is embedded into some middle frequency bands once performing a DCT. Insertion and extraction processes depending on a back-propagation neural network architecture (BPNN). Furthermore, the choice of frequencies and the block covering the watermark contingent on an earliest study of the effect of MP3 coding at different rates on the sound signal. Experimentations display that the suggested scheme presents good robustness and audio quality results. We consider also in the same paper [30], the adaptation of the proposed scheme in video watermarking approach which is different from our previous technique [31] focusing on only the video frames without considering the audio channel. In fact, in [30], we have adjusted the MP3 study to video watermarking scheme with an earliest study of the MPEG video coding. Once more, we achieve the copyright protection purposes and we ameliorate the robustness criteria of the video watermarking technique.

In the same application perspective, we have implemented in [32], a robust and blind image watermarking technique in the frequency domain. In this paper, the algorithm is resistant to diverse types of attacks such as geometric transformations, communal signal processing, standard JPEG compression and even to double Stirmark attacks. This significant robustness is due to the insertion frequency domain, to the choice of the appropriate blocks depending on a preliminary study between the original and the compressed-decompressed image and to the use of the Arnold transformation [32] scrambling the watermark and ameliorating then the security level.

We will describe in section III.A a novel audio watermarking scheme for copyright protection application based on preliminary attacks and frequency masking studies and on spectral envelope estimation of basic and sensitive audio signals.

7) DIGITAL RIGHTS MANAGEMENT

They cover mechanisms used by content publishers and rights holders to inflict access-licensing terms. They concern principally DRM for relational data, precisely database watermarking techniques [33].

8) TRACEABILITY

Digital watermarking is exploited to trace the sender of the digital document copy [34]. The idea is to use a particular mark for each copy. If there is an unlawful copy in the market, we can identify effortlessly the person who distributed it illegally [35].

We were previously interested in the traceability as watermarking application in a first developed technique as described in [36]. In this paper, we have remarked that the tracing operation is frequently constrained by the absence of evidence about the number of colluders and also the collusion channel. Certainly, the Tardos decoding is invariant regardless the type of collusion, that can be reflected its accusation performance. Thus, we proposed to use a MAP-based estimation strategy, increasing the Tardos decoding step and assuring a respectable estimation results. The proposed idea takes the benefit of operating in hierarchical context to deliver a more succinct and exact accusation decision in a short time.

In a second developed tracing scheme detailed in [37], we proposed a confident fingerprinting approach based on a two-stage tracing strategy combining the Boneh Shaw with replication scheme and the Tardos codes. This scheme is applied to a multilevel hierarchical fingerprint hidden by using a DCT-based audio watermarking algorithm [38]. By taking the advantage of grouping users and applying a weight-based tracing mechanism, the suggested fingerprinting technique diminishes well the computational costs of the tracing time and delivers a suitable solution reducing considerably the users' recovery space and performing respectable robustness.

9) INTEGRITY VERIFICATION

The signature is hidden in the original document, and is used more lately to check if its content has been modified or not. In fact, we embed a mark in the document so that if we remove a part of it, portion of the signature will also be removed and this will prevent the correct detection. If the watermark is not detected, we can conclude that the document was altered [39, 40]. In this type of application, we have already developed a semi-fragile audio watermarking technique for MP3-encoded files using Huffman data described in [41] in the compressed domain. The mark is inserted in MP3 bit streams. The algorithm uses mainly big values region and recompression calibration of Huffman data to embed secret information. Experimentations prove the inaudibility of the suggested method and its robustness to several attacks.

We have also treated recently the integrity control application by an image watermarking scheme described in [11]. It fact, this scheme extracts features from the original digital image to generate a watermark. In order to resist rotation and cropping attacks, the technique adopts Speeded-Up Robust Feature [42] to localize invariant key points. Experimentations prove that our scheme gives a high level of invisibility and robustness to standard JPEG compression and unique/double Stirmark attacks and that the integrity is successfully achieved.

10) CONTROL OF COPY AND PLAYBACK

It is probable for playback devices to react to hidden signals. Thus, if the proprietor desires to implement such a system where the duplication recording is forbidden, then manufactured recorder need to embrace mark detection circuitry [43-45].

11) LOCATING DIGITAL CONTENT ONLINE

Digital contents are uploaded on the internet in a large volume designed for research, distribution, and communication tenacity. It has also become a prevalent platform for sales. Thus, the proprietor identification becomes imperative which is conceivable with the help of watermarking [46].

12) FORENSICS

This technique enhances the possibility for the proprietor to detect and respond to the abuse of its possessions. It is exploited not only to gather the proof for criminal, but also to enforce the contractual usage agreement between the proprietor and the individuals with whose it shares its digital content [47].

13) MEDICAL USAGES

Using the approach of visible watermarking, patient details can be reproduced on the Magnetic Resonance and imaging (MRI) and the X-ray scans reports. If the reports of diverse patients are mixed, then the incorrect diagnosis of a malady for a patient based on unknown report may conduct to an unfavorable treatment. Consequently, embedding in a report patient name and date for example could decrease the possibility of maltreatment and increase the security and confidentiality of the patient [48].

B. NON-SECURITY WATERMARKING APPLICATIONS

In the non-security watermarking applications, the robustness to intentional attacks is not necessary and the watermark should generally contain a great capacity information's and must be extracted with a blind detection scheme. Among these applications, we find:

1) BROADCAST VERIFICATION

It aims to compile statistics on the use of the digital content. In radio broadcasts, advertisers commonly want to guarantee that their announcements were suitably distributed according to the number of times specified in the contract. Therefore, a watermark is hidden in each advertisement. It permits, for example, to recognize in which radio the audio signal was broadcasted, how often and even at what time [49].

2) MUSICAL EXTRACTS SEPARATION

A set of information with certain characteristics can be extracted from audio files. This information is hidden inaudibly by watermarking in the mixture of audio sounds. After extraction this embedded watermark, the recovered information permits the separation of the original music signals. [50].

3) INCREASING TELEVISION PROGRAMS INTELLIGIBILITY

It wishes to replace in real time the teletext display by implanting a cloned into the television programs. This will allow deaf and hard hearing people to develop their understanding thanks to the movement of a face and hands reproducing by the Cued Speech [51].

4) SOUND ANNOTATION

It can be used to transfer a label to aid signals indexing. The embedded information can include meta-data describing the signal content or information about a target application [52]. In this context, we have earlier introduced in [53] a watermarking scheme performing a multimodal video characterization and summarization. So, audiovisual features are inserted as the watermark. Using the descriptors enclosed in the mark, key moments within a video, characterized generally by high loudness or high motion, can be recovered just by extracting the equivalent signature. Similarly, narrative video sequence, commonly known by low or medium motion loudness and activity, can be designated using the used watermark. Besides, we can browse within the digital video and we can extract scene with particular properties such as natural or artificial scene, night or day scene.

We will describe in section III.B a new audio watermarking technique for content characterization based on deep learning audio classification scheme.

5) MOBILE USAGES

Digital watermarks offer a marvelous opportunity for marketers looking for new behaviors to engage consumers with rich media experiences on their phones. The watermarks can be easily hidden into all forms of media document, including packaging, newspapers, posters, brochures, etc. [54]. Once an application is downloaded to the smart phone, we simply launch it, hold it parallel around 6″ from the printed content and the smart phone will directly detect the watermark and link then the customer to premium content online. The watermark is accorded to an URL in a backend database that is consequently reverted to the consumers' smart phone.

6) MEASUREMENT OF AUDIENCE

Services of audience measurement must nowadays report more precisely and consistently from several channels. Watermarking hides a single identifier into digital content while being distributed or prior to dissemination, making it and corresponding broadcasters quickly identifiable. The watermark covers evidence about the channel that transmits the program, its exposure time and its media content identifier. Audiometers mounted in panelists' homes read the data, gather the information and conduct them to a central database for treating and perfect reporting daily [55,56].

C. WATERMARKING SENSITIVE DATA

As more communication and collaboration occurs in the digital space, the requisite for maintaining data and document integrity is rising. Thus, businesses try to increase their cloud security budget. As an added layer of security, they often choose to watermark their digital documents when shared internally or externally. Watermarking aids deter recipients from data exfiltration activity, guaranteeing that sensitive information such as contracts, budgets, confidential communication or manuscripts, health records, stays private and compliant during its lifecycle, so collaboration will be achieved with confidence.

For example, in the teleradiology context, privacy and security of sensitive information has become a serious issue [57, 58]. Teleradiology has been understood extensively to be an eHealth service ended through remote diffusion of the radiology information and images above electronic networks, and the interpretation of the transferred images for diagnosis purposes. This radiology data, chiefly Electronic Personal Health Information (EPHI), are exposed to potential altering with severe complications, since they are very sensitive. Such information necessitates protection with integrity and great confidentiality.

Another example concerning identity cards, which also are very sensitive and must be highly concerned. In fact, if the National ID card undergoes attacks like forged identity and counterfeit cards or falsification of content, that will affect citizens and locate the issuing government in excruciating situation. Sensitive National identity card should have then visible and invisible digital watermarking with inserted secrete text information [59].

A third example of sensitive data to be protected is the Arabic Quran recitation requiring [60]. A specific mechanism based on watermarking scheme must execute a number of functional stages avoiding then the distortion of the Quranic signal and addressing successfully its sensitivity. Sensitive Holy Quran in image format is also studying in [11, 32, 61] to detect any manipulation on the Quranic sensitive content and to preserve its content's integrity. Besides, a related diacritical watermarking scheme to secure sensitive Quran Arabic in digital text format is proposed in [62]. Due to sensitivity of Holy Quran, diacritics play an essential role in the sense of the specific verse. Henceforth, acquiring letters with certain diacritics will conserve the original sense of Quranic verses in case of illicit tampering attempt.

Preservation the sensitive nature of certain data needs special digital watermarking algorithms, which are defies that need to be worked on.

III. PROPOSED WATERMARKING SCHEMES

A. WATERMARKING TECHNIQUE FOR COPYRIGHT PROTECTION APPLICATION

We begin by discussing some previous audio watermarking schemes promising the copyright protection of digital audio signals. After that, we introduce our contribution in this type of audio watermarking applications.

1) WATERMARKING TECHNIQUES RELATED TO COPYRIGHT PROTECTION APPLICATION

Copyright protection applications of digital content has become an essential issue. Digital watermarking techniques has received excessive deal of attention to elucidate this problem. This paragraph presents the review of same papers, which mainly focus on copyright protection context.

Paper in [63] presented a 3-level lifting wavelet transform (LWT)-based framework for audio watermarking. To increase applicability, the robust signature including proprietary information, synchronization code, and frame-related data was mainly hidden in the approximation subband by using perceptual-based rational dither modulation (RDM) with adaptive quantization index modulation (AQIM). Experiment results indicated that the hidden robust signature can withstand usually faced attacks. In addition, the system was resistant to cropping and replacement attacks and caused only slight degradation.

A new audio watermarking technique with good robustness was discussed in [64] by discovering the multi-resolution characteristic of the Discrete Wavelet Transform (DWT) with the energy compaction capability of the Discrete Cosine Transform (DCT). The watermark is embedded by slightly altering some frequencies of the audio signal. The audio fragments are segmented by DWT to obtain numerous groups of wavelet coefficients with several frequency bands, and the fourth level detail coefficient is then selected to be alienated into the former packet and the latter one, which are effected for DCT to obtain two sets of transform domain coefficients correspondingly. Lastly, the average amplitudes of the two sets are changed to hide a binary image. The watermark detection is blind. Experimental results endorsed that the suggested algorithm had good inaudibility, large capacity and good robustness when fighting to various attacks.

Another paper in [65] presented an audio watermark technique in DWT domain based on mean-quantization using planar and binary image as signature, and encrypting it with chaos sequence. In this scheme, the audio file is segmented using suitable wavelet basis. Low-frequency coefficients are designated to hide watermark using mean-quantization algorithm. The watermark can be detected without the original audio file. Experimentations showed that compared with known prior quantization watermark embedding schemes, the suggested technique was robust to different attacks.

In [66], a blind and adaptive audio watermarking technique was suggested based on chaotic encryption in DCT and DWT hybrid domain. The encrypted mark can be hidden into the audio signal according to the special insertion rules. The hidden depth of each segment is controlled by the overall average amplitude to efficiently increase the inaudibility and the robustness. The signature is encrypted by a chaotic sequence to enhance the security of the watermark. Experimental tests displayed that the suggested technique had higher capacity, good inaudibility, larger security, and good robustness when opposing signal-processing attacks.

A blind technique proposed in [67], jointly exploring in DWT the auditory masking properties and the rational dither modulation (RDM). The insertion of binary information is assured by modulating coefficient vectors in the 5th-level approximation subband. The robustness and capacity of the suggested scheme are controllable by changing vector dimensions, while the inaudibilty is guaranteed by constraining quantization noise under the auditory masking threshold. Besides, the periodic characteristic inbred in the RDM formulation can be exploited to re-ensure synchronization for truthful watermark extraction. Experimentations displayed that the DWT–RDM technique furnished a near-zero objective difference grade even when the SNR sustained at a level near 20 dB. In most attacks, the bit error rates BERs were suitably low as associated to other lately developed methods with littler capacities.

In [68], this paper proposed a technique which inserts the watermark into the maximal coefficient in DCT of a moving average sequence. In fact, signal processing operations generate noise that usually modifies the high frequencies of an audio file. Thus, hiding watermark by regulating low-frequency coefficient can enhance the robustness of a watermark algorithm. Moving Average sequence is a low-frequency feature of an audio file. Subjective and objective tests divulged that the suggested watermarking technique preserves highly audio quality, and at the same time, the scheme is highly resistant to most known digital signal processing manipulations.

We introduce in the following the new proposed watermarking technique for copyright protection application.

2) INTRODUCTION OF THE WATERMARKING TECHNIQUE FOR COPYRIGHT PROTECTION APPLICATION

In this section, we introduce an enhanced approach of our previous audio watermarking technique called DCT-NN [2] based on Neural Network NN architecture. The new watermarking scheme presents a new approach to address the challenges associated with copyright protection of basic and sensitive audio data like Quranic files but can also be extended to assure their content integrity and tamper detection. In this approach, we insert the watermark after performing DCT transform into middle frequency bands. To improve robustness and security while maintaining good inaudibility results, we exploited BPNN architecture in the embedding and extraction processes [30]. The basic idea is to establish the relationship between frequency samples around a central sample by using the BPNN model. In fact, for a selected transformed sample I(x), the NN is trained with its 8 neighbors as input vector and the value of the sample as output. The used BPNN architecture contains three layers: the input layer with eight neurons, the hidden layer with nine neurons, and the output layer with a single neuron. After performing the frame division of the original audio signal, the DCT transform is applied to the resulted frames. Next, each transformed frame is divided into nine samples forming a block as shown in figure 3. The center sample of the block is the output and the neighbor's samples are the input. We proceed finally to NN training until a definite goal or a specified maximum number of iterations is reached. When the BPNN training is completed, a set of synaptic weights (wi) characterizing the behavior of the trained network can be obtained and used in the BPNN simulation of the embedding and the extraction processes.

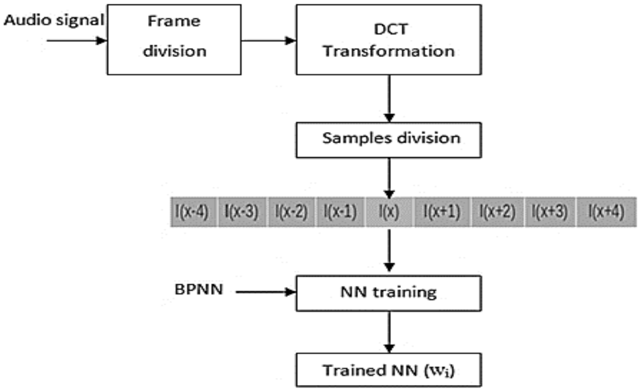


FIGURE 3. BPNN training process

The originality of this new scheme is due to the exploitation of the frequency perceptual masking of the Human Psychoacoustic Model HPM [69] associated to the Linear Predictive Coding LPC [70] spectral envelope estimation of the digital audio file. In fact, after studying the HPM, we examined the masking threshold curve Ltg [69] and we compared it with the LPC envelope to hide properly and imperceptibly the watermark under this curve. Another specificity of the scheme, is the totally blindness detection process unlike previous schemes [2, 30], as neither original audio signal nor secure key are saved and transmitted to the receiver. In fact, frame positions and correspondent indexes of insertion are recalculated in the detection process, which guarantee its blindness. Experimental results indicated that the exploitation of perceptual masking with the spectral envelope consideration in the frequency domain is very interesting with very good robustness results.

3) PRELIMINARY STUDY

Preliminary study of original WAVE signals is performed before the watermarking process.

The result of this preliminary study is a classification of the Stirmark attacks before watermarking permitting us to choose the adequate attacks that are suitable to the copyright protection context. The robustness of our scheme is evaluated again the chosen attacks after the watermarking process. So, different MATLAB simulations are achieved. Table 1 displays a selection of studied standard music signals and sensitive Quranic audio files. All signals have sampling rates of 44.1 KHz, number of bits per sample equal to 16bps and duration around of 20 s.

TABLE 1
SELECTED ORIGINAL AUDIO FILES

Name	Description
Standard Musical files	
Tunisia	Rhythmic music
Svega	Female song voice
Sensitive Quranic files	
Track 01	Alfatiha
Track 02	Extract of Elbakara
Track 03	Alfil
Track 04	Alnasr

To justify the chosen attacks that we will test for measuring robustness property of our proposed watermarking technique, we are based on one of the International Federation of Phonographic Industry IFPI exigencies [30] which inflicts that the watermarking algorithm must avoid unauthorized removal of the hidden watermark unless the audio signal quality becomes very humble. Therefore, we have applied in the preliminary study, all the Stirmark attacks to the original audio signals to discern audio quality after attacks and to discard then attacks that corrupt remarkably the audio quality. In fact, if an audio signal is extremely damaged, robustness will not certainly guaranteed. We have computed the Signal to noise ratio SNR [2] values, which are measured in decibels, between the original audio signals and the corresponding attacked ones. Besides, to well verify the quality audio, we have achieved the Subjective Difference Grade SDG tests based on Recommendation UIT-R BS.1116 [64] and we have obtained their values with their descriptions.

The preliminary studies are exposed in the tables 2, 3, 4 and 5.

TABLE 2
IMPERCEPTIBILITY STIRMARK ATTACK PART-1 TESTS

	Stirmark attacks	Tunisia.wav			Svega.wav		
		SNR	SDG/Description		SNR	SDG/Description	
1	Exchange	15.37	-1	Perceptible but not annoying	14.71	0	Imperceptible
2	Extrastereo 30	67.88	0	Imperceptible	60.06	0	Imperceptible
3	Extrastereo 50	76.09	0	Imperceptible	59.91	0	Imperceptible
4	Extrastereo 70	80.93	0	Imperceptible	59.77	0	Imperceptible
5	Invert	-6.02	0	Imperceptible	-6.02	0	Imperceptible
6	Fft_invert	-6.02	0	Imperceptible	-6.02	0	Imperceptible
7	Fft_real reverse	31.57	0	Imperceptible	47.01	0	Imperceptible
8	Lsbzero	63.11	0	Imperceptible	66.29	0	Imperceptible
9	Normalize	17.73	0	Imperceptible	17.44	-1	Perceptible but not annoying
10	Rc_highpass	7.48	0	Imperceptible	7.16	0	Imperceptible
11	Rc_lowpass	24.15	0	Imperceptible	24.24	0	Imperceptible
12	Smooth	29.67	0	Imperceptible	22.55	0	Imperceptible
13	Smooth2	28.16	0	Imperceptible	23.92	0	Imperceptible

TABLE 3
IMPERCEPTIBILITY STIRMARK ATTACK PART-1-BIS TESTS

Stirmark attacks		Tunisia.wav			Svega.wav		
		SNR	SDG/Description		SNR	SDG/Description	
14	Stat1	21.39	0	Imperceptible	20.73	0	Imperceptible
15	Stat2	35.32	0	Imperceptible	29.68	0	Imperceptible
16	Re_sample 44.1_32_44.1	62.81	0	Imperceptible	44.57	0	Imperceptible
17	Re_sample 44.1_22.5_44 .1	43.29	0	Imperceptible	27.95	0	Imperceptible
18	AddBrumm 100	37.05	0	Imperceptible	29.08	0	Imperceptible
19	AddBrumm 1100	16.18	0	Imperceptible	8.21	0	Imperceptible
20	AddBrumm 2100	10.56	0	Imperceptible	2.59	0	Imperceptible
21	Addnoise 100	39.40	0	Imperceptible	31.44	0	Imperceptible
22	Addnoise 300	29.82	0	Imperceptible	21.85	-1	Perceptible but not annoying
23	Conversion 16_8_16	30.63	0	Imperceptible	22.58	-1	Perceptible but not annoying
24	Cut_replace samples_1	91.54	0	Imperceptible	43.81	0	Imperceptible
25	Cut_replace samples _10	81.64	0	Imperceptible	43.82	0	Imperceptible
26	Cut_replace samples_20	78.56	0	Imperceptible	43.82	0	Imperceptible

- Studying Stirmark attack part-1 (attacks from 1 to 26) in tables 2 and 3

When applying these attacks, we do not perceive a degradation of the audio quality. All subjective results are often "imperceptible" with an SDG=0 and sometimes "Perceptible but not annoying" with an SDG=-1. For the SNR values, they are positive values except for the attacks "invert" (number 5) and "fft_invert" (number 6) of table 2. In effect, for the "invert" attack, the principle is to replace each sample value by its opposite and for the "Fft_invert" attack, the principle is to invert both the real and the imaginary in the frequency domain of the sample values. Therefore, it is obviously to get negative SNR values while having no degradation in audio quality (SDG=0, description="impercebtible"). For the attacks from 26 to 29, they are audio manipulations that do not exist in the Stirmark attacks. Conversion 16_8_16 (number 23) changes the number of bits per sample from 16 to 8 bits and vice versa. Cut_replace_samples_1, Cut_replace_samples_10 and Cut_replace_samples_20 (numbers 24, 25, 26) are combinations of two attacks "cutsamples" (number 51) and "copysamples" (number 52) described in the table 5. For example, Cut_replace_samples_20 removes twenty samples every 1000 samples and replaces them by another twenty samples.

After examining the inaudibility Stirmark attack part-1 studies from tables 2 and 3, we notice that these attacks do not affect the audio quality of the audio files. Thus, we will consider them in our watermarking robustness experiments and we anticipate that we will obtain good robustness results against these attacks.

TABLE 4
IMPERCEPTIBILITY STIRMARK ATTACK PART-2 TESTS

Stirmark attacks		Tunisia.wav			Svega.wav		
		SNR	SDG/Description		SNR	SDG/Description	
27	AddBrumm 3100	7.18	0	Imperceptible	-0.79	-1	Perceptible but not annoying
28	AddBrumm 4100	4.75	0	Imperceptible	-3.22	-2	Slightly Annoying
29	AddBrumm 5100	2.86	0	Imperceptible	-5.11	-3	Annoying
30	AddBrumm 6100	1.30	0	Imperceptible	-6.67	-3	Annoying
31	AddBrumm 7100	-0.01	-1	Perceptible but not annoying	-7.99	-3	Annoying
32	AddBrumm 8100	-1.16	-1	Perceptible but not annoying	-9.13	-3	Annoying
33	AddBrumm 9100	-2.17	-2	Slightly Annoying	-10.14	-3	Annoying
34	AddBrumm 10100	- 3.08	-2	Slightly Annoying	-11.05	-3	Annoying
35	Addnoise 500	25.38	0	Imperceptible	17.40	-2	Slightly Annoying
36	Addnoise 700	22.46	0	Imperceptible	14.48	-4	Very Annoying
37	Addnoise 900	20.27	0	Imperceptible	12.29	-4	Very Annoying
38	Amplify	6.01	-3	Annoying	6.02	-2	Slightly Annoying
39	Compressor	21.46	-2	Slightly Annoying	60.21	0	Imperceptible
40	Dynnoise	19.32	0	Imperceptible	19.31	-1	Perceptible but not annoying
41	Fft_hlpass	11.81	0	Imperceptible	17.44	-1	Perceptible but not annoying
42	Zerocross	25.88	0	Imperceptible	15.87	-3	Annoying

- Studying Stirmark attack part-2 (attacks from 27 to 42) in table 4

When applying these attacks, we observe that there are distinguished irregularities in the results:

Irregularity type 1: at this point, we observe that the results vary from one signal to another. In effect, we can find for the same attack "imperceptible", "slightly annoying", "annoying" and "very annoying" as decision of the subjective results.

Irregularity type 2: here, we perceive for the same signal and the same attack that the results between the objective test SNR and the subjective test SDG are not equivalent. For example, for the audio file "svega.wav" and the attack "addnoise_500", we obtain 14.48 as SNR but with "very annoying" as decision of the subjective test. However, for the audio signal "Tunisia.wav" and the attack "add_brumn_8100", we find "Perceptible but not annoying" as decision of the subjective test with a bad SNR value equals to -1.16.

For the Stirmark attacks part-2 presented in table 4, we cannot expect the watermarking robustness results after applying them to the watermarked audio signal as we cannot make a global decision if they corrupt or not the audio quality.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3145950, IEEE Access

These attacks will be deliberated in our watermarking robustness tests.

- Studying Stirmark attack part-3 (attacks from 43 to 52) in table 5

We perceive without doubt a significant degradation of the audio quality. In fact, all subjective results are all time "Very Annoying" with an SDG=-4 for all original audio files. For the SNR values, they are bad with lower values except for the attack "addsinus" (number 43). Besides, the attacks from 49 to 52 are the worst attacks that affect remarkably the audio quality. In effect, in addition to the fact that the resulted subjective decisions are almost "Very Annoying" with an SDG=-4, it is not possible to calculate the SNR with these attacks as the obtained attacked audio files are very different from the original (they do not have the same dimensions). As our proposed audio watermarking technique is typically used for copyright protection application, we conclude that it is not interesting to take into account the attacks of table 5, in the robustness tests.

TABLE 5
IMPERCEPTIBILITY STIRMARK ATTACK PART-3 TESTS

Stirmark attacks		Tunisia.wav			Svega.wav		
		SNR	SDG/ Description		SNR	SDG/ Description	
43	Addsinus	14.73	-4	Very Annoying	6.75	-4	Very Annoying
44	Echo	3.14	-4	Very Annoying	2.98	-4	Very Annoying
45	Flippsample	0.45	-4	Very Annoying	0.64	-4	Very Annoying
46	Fft_stat1	1.23	-4	Very Annoying	1.63	-4	Very Annoying
47	Addfftnoise	0.01	-4	Very Annoying	9.27e-004	-4	Very Annoying
48	Voiceremove	-4.61e-006	-4	Very Annoying	-3.85e-006	-4	Very Annoying
49	ZeroLength	X	-4	Very Annoying	X	-4	Very Annoying
50	ZeroRemove	X	-4	Very Annoying	X	-4	Very Annoying
51	Cutsamples	X	-4	Very Annoying	X	-4	Very Annoying
52	Copysamples	X	-4	Very Annoying	X	-4	Very Annoying

In fact, applying these attacks to the watermarked audio file noticeably corrupts the audio quality, and then the attacked watermarked file will not be exploited. Despite these facts, and to observe the behavior of our watermarking approach against these malevolent attacks, we decide to test the robustness against three selected attacks of table 5 which are "addsinus" (number 43), "echo" (number 44) and "flippsample" (number 45). This choice is for the reason that it is possible for a pirate to apply them to remove the watermark without realizing that it will damage the auditory quality of the attacked watermarked signal. Furthermore, we have also thought of combining two attacks and perceiving their effects on the watermark. Since the attacks "cutsamples" (number 51) and "copysamples" (number 52) significantly destroy the audio quality if they are applied each one alone, we decide to combine them. We first remove one (or ten) (or twenty) samples every 1000 samples (the "cutsamples" attack) and then we replace them (the "copysamples" attack) by one (or ten) (or twenty) corresponding samples of the original audio file. The obtained attacks are "Cut_replace_samples_1", "Cut_replace_samples_10" and "Cut_replace_samples_20". We categorized the obtained attacks in the table 3 (numbers 24 to 26) as they always present very high SNR and subjective results "imperceptible" with an SDG=0.

We explicate in the following paragraphs our proposed audio watermarking technique for copyright protection application.

4) EXPLOITATION OF HPM WITH LPC ESTIMATION IN A NEW AUDIO WATERMARKING TECHNIQUE FOR SECURITY COPYRIGHT PROTECTION APPLICATION

The MPEG audio standard [69] encodes audio file by eliminating the acoustically irrelevant portions of the audio data. In reality, it benefits from the human auditory system's incapability to perceive quantization noise beneath auditory masking conditions. The HPM calculates the quantitative estimation of the basic limit of indiscernible audio signal compression. This limit is the masking threshold curve Ltg deliberated after performing HPM seventh steps. HPM imposes that to have an imperceptible quantization noise; this noise should stay below the masking threshold curve. We have tried to make an analogy between compression and watermarking. Since the quantization noise resulted from the MPEG audio compression is inaudible when it is under the Ltg, we have anticipated then that the noise caused by the watermark insertion will be also inaudible if it is under the Ltg.

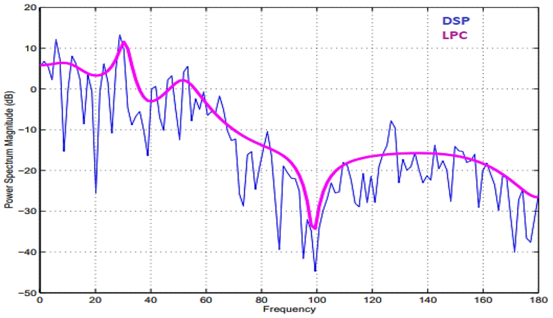


FIGURE 4. The smoothing aspect of LPC curve with minimum variations vs PSD curve

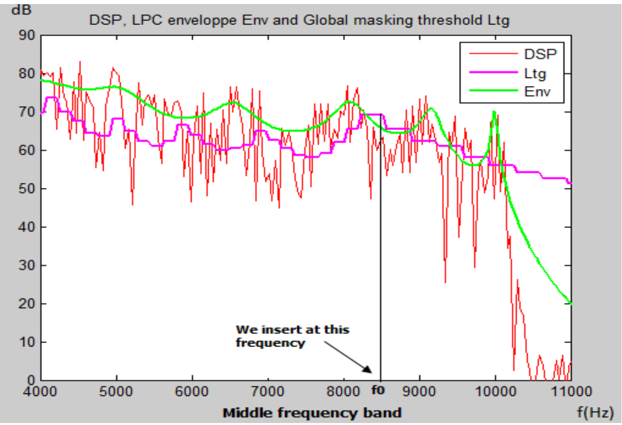


FIGURE 5. The estimated envelope Env compared with Ltg

The new proposed audio watermarking technique for copyright protection applications DCT-NN-HPM followed these steps:

- The first seventh stages of the Human psychoacoustic model 1 HPM1 [69] are performed to obtain the masking threshold curve "Ltg". The first step of the HPM1 computes the PSD of a 512 audio frame. These frames are overlapped with 128 samples as joint part.
- The non-overlapped frames noted "sframe" in temporal domain are given also after frame division of the original audio signal. Each non-overlapped frame "sframe" is of 384 samples size. These frames will be used later to embed the watermark bit after a 384-DCT transform.
- Getting the DSP from the first HPM1 stage, we calculate its envelope using the LPC envelope estimation "Env". LPC envelope "Env" is chosen instead of the PSD to improve robustness of the scheme as LPC presents, after attacking audio signals, a smooth curve with minimum variations unlike the PSD curve as depicted in figure 4. LPC is universally used for sensitive envelope estimation and offers a smooth representation of the important and delicate sound proprieties. The idea of the LPC estimation is to represent each current audio sample x(n) by a linear combination of its p prior values x(n-p-1) through x(n-1). p is the order of the LPC[70]. Figure 5 displays in the middle frequency band [4 KHz, 11 Khz] the LPC envelope estimation "Env" of the PSD of an elected audio frame and the matching calculated "Ltg". As depicted, f0 can be the adequate frequency where we insert delicately the watermark bit in the sensitive selected frame.
- After localizing the middle frequency MF band in a range of an audio frame depending on the audio signal characteristics, we compute the positive variances in the MF so that the envelope "Env" is under the "Ltg" as following:

❖ For all samples in the MF band of a 512 frame do:
if Ltg > Env then diff_positive=Ltg-Env

❖ We calculate next the maximum difference from the computed positive differences "diff_positive". It is imperative to note that the localized middle frequency band must be significantly narrow so that it will be the same calculated during the detection to ensure resynchronization of the frames and insertion positions.

- Lastly, after accomplishing the three previous steps for all overlapped 512 frames, we obtain N frequencies values where we can embed the watermark. We necessary generate a mapping between the indexes corresponding to these frequencies in the overlapped 512 frames and the indexes of the non-overlapped 384 audio frames "sframe". We hide the watermark bits in the suitable index of the selected "sframe" after converting it to the frequency domain.

We define the embedding and the detection processes of this scheme in the following paragraphs:

- DCT-NN-HPM watermark embedding process

- In the previous DCT-NN audio watermarking scheme, the audio signal is separated into non-overlapping frames of 512 samples and a DCT transform is achieved to each obtained frame. However, in the new DCT-NN-HPM, the provided non-overlapped frames "sframe" from the original audio division are of 384 samples size. Accordingly, the result is a DCT frame of 384 frequency samples size noted "sframe_DCT". The obtained "sframe_DCT" is used after that to cover the watermark bit.
- In the preceding DCT-NN audio watermarking approach, we have chosen to hide the watermark bit in middle frequency band [4 kHz, 11 kHz]. For each frame and after localizing this band, we have explored the sample value the closest to the average value of the middle frequency located band and then we have deducted its position. The sample of the identified position covered the watermark bit. However, the research of the position of insertion in the new DCT-NN-HPM approach is different. In fact, after localizing a narrow middle frequency band depending on the audio signal characteristics, we have computed the positive differences in this band so that the LPC envelope is below the Ltg. Afterward; we have computed the maximum difference from the deliberated positive differences. The frequency sample corresponding to this maximum difference covered the watermark bit. The watermark insertion steps of the new DCT-NN-HPM are illustrated in the Figure 6.

- DCT-NN-HPM watermark extraction process

The DCT-NN-HPM detection process is displayed in Figure 7. The searched frames and the positions of insertion constitute the proposed audio watermarking key. It is essential to note that this key is not transferred secretly to the receiver but it is recalculated in the detection process which implicate the totally detection blindness of the new technique in contrary to the previous technique. In fact, searched frames and positions of insertion from an adequate narrow middle frequency band are the result of applying the first seventh steps of the HPM1 on the watermarked audio file. This stage is very significant in the detection process since it assures the re-synchronization of de-synchronized frames and correspondent insertion positions in the case of de-synchronizing attacks. Another difference with the extraction process of the DCT-NN scheme is that the watermarked audio file is separated into non-overlapping frames of 384 samples as exhibited in Figure 7. We display in the following paragraphs the experimental results of the DCT-NN-HPM and the comparison tests with DCT-NN and other audio watermarking schemes.

5) INAUDIBILITY AND ROBUSTNESS RESULTS OF THE SECURITY WATERMARKING APPLICATION

To test compression robustness, we used standard lame Audio Encoder [30]. Besides, for other audio operations, we used the standard StirMark Benchmark for Audio (SMBA) tool with default parameters [71] and Audacity 2.3.3.

We used as watermark a binary image of size 32 × 32.

Two common robustness evaluation metrics utilized in the literature are the normalized cross-correlation NC [2, 30] and the Bit Error Rate BER [2, 72, 73]. They assess the similarity between the extracted watermark and the inserted one. More NC is near to 1, more extracted watermark is similar to the embedded watermark. In the contrary, more BER is near to 0, more extracted watermark is similar to the hidden one.

In our tests, we assume that the watermark that is a binary logo of size 32 × 32, is existent if the calculated correlation exceeds 0.7 as chosen threshold value. In fact, if NC surpasses this threshold, the extracted watermark is visibly similar to the hidden watermark. Moreover, we consider that the watermark is decorously extracted, if the computed Bit Error Rate value is less than 0.3. In effect, if BER is under this threshold, the detected watermark is perceptibly comparable to the embedded one.

As we know, the most famous removal attack is lossy compression. The common standard lossy compression for audio signal is the MPEG 1 Audio Layer III MP3 that is regularly used by audio consumer storage. Different bit rates are used in the MP3 standard. 128 Kbps bit rate is the most usually used [74] at a compression ratio of 11:1, guaranteeing generally adequate sound quality. We test robustness of the actual proposed watermarking approach using three MP3 compression rates (128, 96 and 64Kbps). This chosen bit rates are the most frequently used rates in prior audio watermarking techniques [19, 41, 63, 64, 67,75, 76, 77].

- Inaudibility results

Figure 8 shows the inaudibility results of the DCT-NN-HPM scheme.

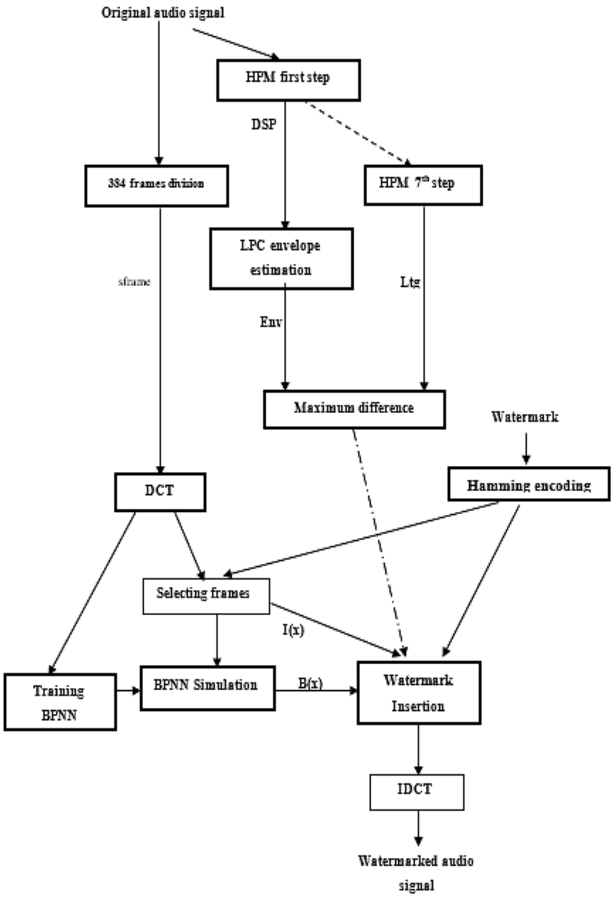


FIGURE 6. DCT-NN-HPM insertion process for security copyright protection application

Due to the exploitation of the frequency perceptual masking related to the LPC estimation of the digital audio signal, obtained SNR values are between 39 dB and 52 dB and are significantly higher than the designed value by the IFPI (20 dB).

- MP3 robustness results

Figure 9 exhibits the MP3 robustness results. For all audio signals, we achieve very good MP3 robustness results (even, with 64Kbps as compression rate, we have all the time NC values greater than 0.87).

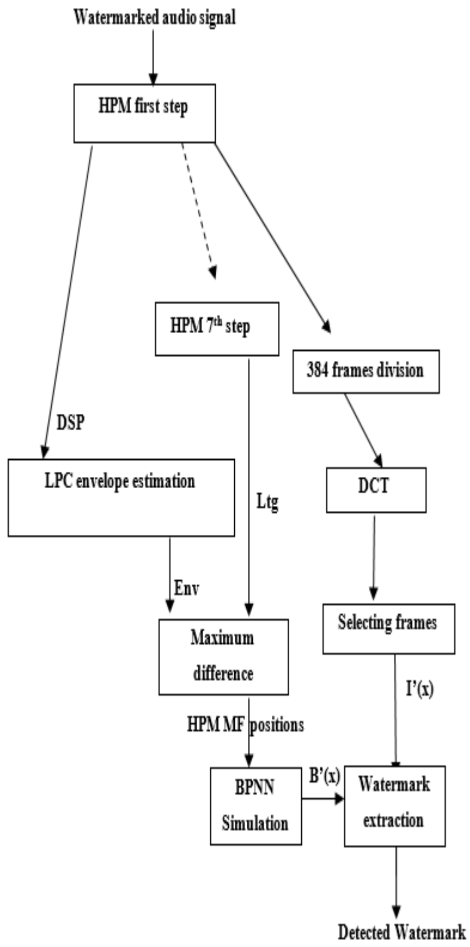


FIGURE 7. The DCT-NN-HPM extraction process for security copyright protection application

The DCT-NN-HPM technique resists truly to the MP3 compression attack even with very damaging bitrates. We realize than, that using the HPM in the frequency domain assures not only perfect inaudibility but also good robustness to MP3 compression.

- Stirmark attacks part-1 robustness results

Figure 10 presents the Stirmark attacks part-1 robustness results. We deduce that the DCT-NN-HPM scheme has good robustness results excepting the invert/fft_invert attacks.

- Stirmark attacks part-2 robustness tests

Figure 11 displays the DCT-NN-HPM based Stirmark attack part-2 tests. We deduce that exploiting the HPM in the frequency domain has noticeably providing good Stirmark attack part-2 robustness results.

- Stirmark attacks part-3 robustness results

Figure 12 exhibits the DCT-NN-HPM based stirmark attack part-3 tests. We obtain satisfying robustness results in spite of the damaging perceptive effects of these types of attacks specially for sensitive Quranic audio signals (NC >0.83)

Lastly, we conclude that experimental results have revealed that the exploitation of frequency perceptual masking studied in HPM with the spectral envelope estimation in the frequency domain are very interesting with very good inaudibility and robustness results.

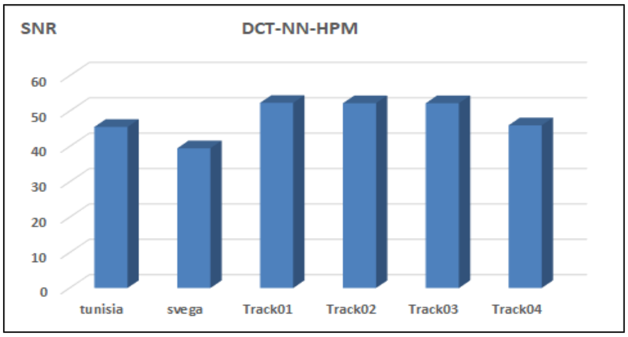


FIGURE 8. Inaudibility results

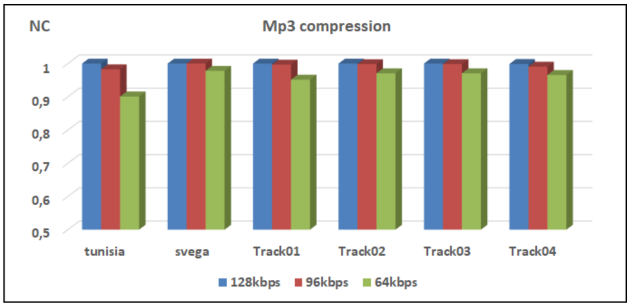


FIGURE 9. MP3 robustness results of the DCT-NN-HPM

6) INAUDIBILITY AND ROBUSTNESS COMPARISON WITH OTHERS

In this section, we exhibit comparison results with our previous scheme DCT-NN [2] and other published audio watermarking schemes by computing the BER, NC and SNR averages of different marks and audio files for all compared schemes.

- Inaudibility comparison with others

Exploring the table 6, we observe that the DCT-NN-HPM approach is the most efficacious audio watermarking scheme in terms of inaudibility. Moreover, our previous scheme [2] and the technique in [78] assure also good imperceptibility results.

- MP3 compression comparison with others

We compare the robustness to MP3 attack of the introduced audio watermarking technique with others. Results are presented in tables 7 and 8. “X” means that the equivalent technique does not treat the indicated attack. When examining the compression results, we notice that our suggested technique DCT-NN-HPM has the best results while using BER or NC metrics when considering the three compression bitrates. This observation proves the effectiveness of integrating the HPM masking study in the embedding algorithm. Besides, schemes in [64, 66, 67, 68, 77] are also robust to MP3 compression.

- Stirmark attacks comparison results with others

The Stirmark attack results are introduced in tables 9 and 10.

In fact, when examining table 9, showing the comparative Stirmark attacks between our proposed scheme and other ones by using the normalized cross-correlation NC, we notice that our suggested scheme DCT-NN-HPM has the best robustness results since all the NC values are 1 or very closed to 1.

Moreover, if we observe table 10, showing the comparative Stirmark attacks between our suggested technique and existing ones by using the Bit Error Rate BER, we remark that our scheme DCT-NN-HPM has the best robustness results since all the BER values are 0 or very closed to 0, excepting the invert attack.

In addition, techniques in [2, 63, 64, 65] resist well to designed Stirmark attacks.

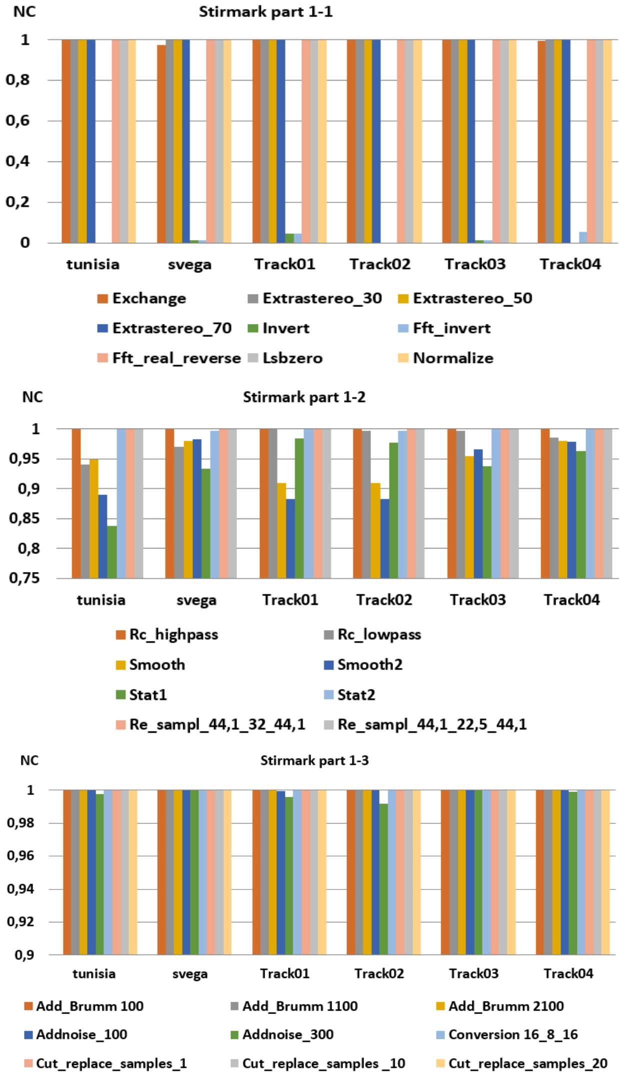


FIGURE 10. Stirmark attack part-1 results of the DCT-NN-HPM

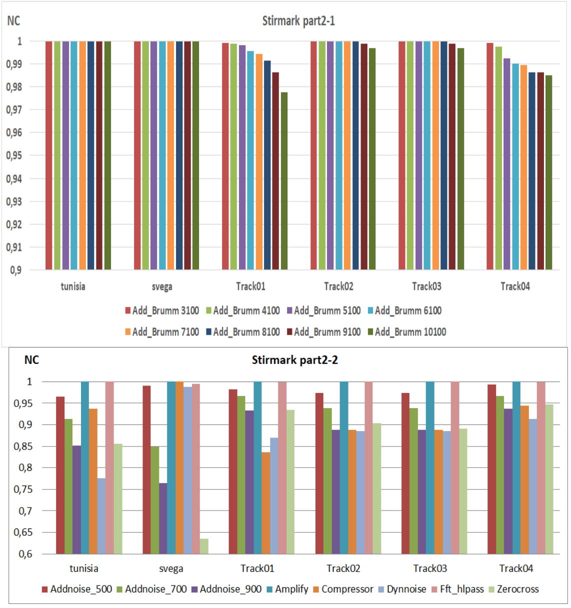


FIGURE 11. Stirmark attack part-2 results of the DCT-NN-HPM

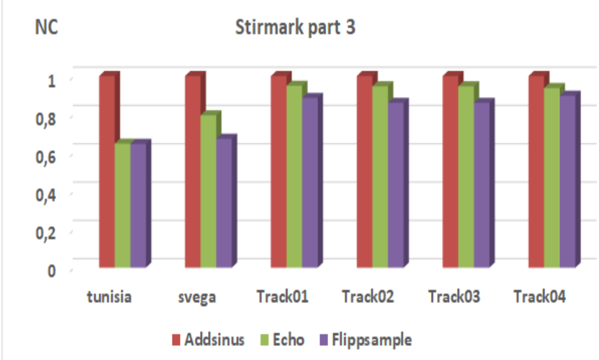


FIGURE 12. Stirmark attack part-3 results of the DCT-NN-HPM

TABLE 6
COMPARATIVE INAUDIBILITY RESULTS OF THE DCT-NN-HPM TECHNIQUE WITH OTHERS

Algorithms	SNR
DCT-NN-HPM	**47,62**
DCT Neural Network architecture [2]	43.52
Support vector regression [75]	27.23
Rational dither modulation with majority voting [63]	28.33
DWT Variable-dimensional vector modulation [67]	20.30
Modifying the Average Amplitude in Transform Domain [64]	23.49
Compressive Sensing [78]	41.54
Wavelet-coefficients quantization [82]	21
Wavelet-coefficients Mean-quantization [65]	37.97
Asymmetric turbo-Hadamard code [83]	29.63
Singular-value decomposition [84]	25.24
Chaotic Encryption in Hybrid Domain [66]	24.58
Spread spectrum [85]	28.59
DC-level shifting [86]	21.24
Echo-data hiding [87]	21.47
Phase-coding [87]	12.2
Frequency-masking [88]	12.87
Empirical Mode Decomposition [89]	25,415
Fast Walsh Hadamard Transform [90]	33.83
Wavelet based technique [91]	32.45
DWT-based rational dither modulation [67]	20.21
Moving Average and DCT [68]	30.93
Air Channel Characteristics [92]	>20

TABLE 7
COMPARATIVE MP3 COMPRESSION (NC) RESULTS OF THE DCT-NN-HPM WITH OTHERS

Algorithms	128 Kbps	96 Kbps	64 Kbps
DCT-NN-HPM	**1**	**1**	**0,95**
DCT Neural Network architecture [19]	1	0.98	0.93
Support-vector regression [71]	0.96	X	X
Modifying the Average Amplitude in Transform Domain [64]	1	X	0.99
Wavelet-coefficients quantizing [82]	X	X	0.84
Wavelet-coefficients Mean-quantization [65]	X	X	0.77
Asymmetric turbo-Hadamard code [83]	1	X	X
Chaotic Encryption in Hybrid Domain [66]	1	X	0.99
Fast Walsh Hadamard Transform [90]	X	X	0.99
Wavelet based technique [91]	0,97	X	X

TABLE 8
COMPARATIVE MP3 COMPRESSION (BER) RESULTS OF THE DCT-NN-HPM WITH OTHERS

Algorithms	128 Kbps	96 Kbps	64 Kbps
DCT-NN-HPM	**0**	**0**	**0.01**
DCT Neural Network architecture [2]	0.0049	0.0098	0.05
Support-vector regression [71]	0.02	X	X
Rational dither modulation with majority voting [63]	0.04	X	10.10
DWT Variable-dimensional vector modulation [67]	0	X	0.009
Modifying the Average Amplitude in Transform Domain [64]	0.01	X	0.08
Wavelet coefficients quantizing [82]	X	X	0.23
Wavelet-coefficients Mean-quantization [65]	X	X	0.29
Singular-value decomposition [84]	0	X	X
Chaotic Encryption in Hybrid Domain [66]	0.01	X	0.06
Fast Walsh Hadamard Transform [90]	X	X	0
DWT-based rational dither modulation[67]	0	X	0.01
Moving Average and DCT [68]	0	X	0.01

TABLE 9
COMPARATIVE STIRMARK ATTACKS (NC) OF THE DCT-NN-HPM WITH OTHERS

Attacks	DCT -NN-HPM	[2]	[71]	[64]	[65]	[66]	[83]
Attack free	**1**	1	1	1	1	1	1
Add noise	**1**	1	X	0.98	X	0.98	0.77
Normalize	**1**	1	X	X	X	X	0.98
Statistical evaluation	**0.99**	0.98	X	X	X	X	0.76
Lsbzero	**1**	1	X	X	X	X	1
Re-sampling 44.1-22.05-44.1	**1**	1	X	1	0.99	X	1
Re-sampling 44.1-32-44.1	**1**	1	0.88	X	X	1	X
LowPass filtering	**0.98**	0.98	0.96	1	0.99	1	X
Convert 16-8-16	**1**	1	1	0.99	0.98	0.99	X

We describe in the following a second audio watermarking scheme for non-security usage focusing on audio content characterization and based on deep learning classification architecture.

TABLE 10
COMPARATIVE STIRMARK ATTACKS (BER)
OF THE DCT-NN-HPM WITH OTHERS

Attacks	DCT -NN-HPM	[2]	[63]	[64]	[65]	[66]	[85]	[92]
Attack free	**0**	0	0	0	0	0	0	0
Echo	**0.09**	0.13	1.49	0.01	X	0.01	0.36	X
Add Brumm	**0**	0	X	X	X		0.01	0.02
AddSinus	**0**	0	X	X	X		0.03	X
Addnoise	**0**	0	0	2.27	X	1.92	0.01	0.05
Amplify	**0**	0	X	0.01	X	0.01	0.51	1.28
Statistical evaluation	**0.04**	0.05	X	X	X	X	0	X
Lsbzero	**0**	0	X	X	X	X	0	X
Invert	**0.57**	0.59	X	X	X	X	0.5	0.02
Re-sampling 44.1-22.05-44.1	**0**	0	0	0.01	0.01	0.01	X	1.38
Re-sampling 44.1- 32.0-44.1	**0**	0	X	X	X	X	X	X
LowPass filtering	**0.01**	0.02	0	0.01	0.01	0.01	X	0.02
Con-version 16-8-16	**0**	0	0	0.14	0.02	0.12	X	X

B. AUDIO WATERMARKING SCHEME FOR DEEP LEARNING BASED AUDIO CONTENT CHARACTERIZATION APPLICATIONS

We begin by debating some prior watermarking techniques related to content characterization applications. After that, we introduce our contribution in this type of audio watermarking applications.

1) WATERMARKING TECHNIQUES RELATED TO CONTENT CHARACTERIZATION APPLICATION

Some prior techniques were debating content characterization by watermarking scheme.

In [93], authors studied two different areas of content-based audio watermarking and recovery using Time-Frequency (TF) parameters. Audio signals are non-stationary and multi-component signals, which involve a series of sinusoids with harmonically allied frequencies. Thus, authors considered the short time Fourier transform (STFT) of the audio file to extract parameters that will be exploited to classify or watermark the signal. Hence, authors suggested a new spread spectrum watermarking algorithm using Instantaneous Mean Frequency (IMF) estimation of the original audio signal and the simultaneous masking to obtain optimal points of watermark insertion. Results confirmed that the watermark was inaudible, statistically unnoticeable and robust to standard signal processing manipulations with BER 0-13%.

In [94], A TF-based audio coding algorithm with new psychoacoustics model, music classification, audio classification, audio fingerprinting, and audio watermarking was introduced to demonstrate the benefits of using time-frequency methods in studying and extracting information from audio files. Authors used IMF estimation of the audio signal and non linear TF signature as mark. They proposed chirp-based watermarking, in which they hidden the linear phase signals as TF signatures. To compensate the BERs in the estimated watermarked audio signal, Hough-Radon transform (HRT) is used as chirp detector in the post-processing process. The technique could correct the error up to BER of 20% and the robustness was acceptable.

In [95], authors used state-of-the-art in frame selection to sugest a new approach to preserve most of the discriminative features of speaker and to safe speech signals by applying speech watermarking method. Thus, linear predictive analysis is exploited for each frame to extract gain, formants and residual errors. Consequently, a frequency weighted function is utilized to quantify formants, and high order correlation with error gain is exploited for weighting the residual errors. Experimentations confirmed an overall (12 %) efficiency in terms of performance, memory and time of frame selection approach for speaker recognition and speech watermarking.

Paper in [96], presented a technique for joining biometric speech authentication and watermarking to assimilate metadata into the authentication process lacking important quality and performance damages. Different audio watermark schemes was introduced to hide metadata as supplementary information into the reference data of biometric speaker recognition. Metadata consisted on auxiliary information about the social, cultural or biological context of the proprietor of the biometric information as well as technical specifics of the sensor. Authors achieved their tests based on a database reserved from 33 subjects and 5 different expressions and a known cepstrum based speaker recognition approach in verification mode. The objective is to accomplish an evaluation of the recognition precision for the selected technique in the context of the gender belongings of the individuals. The first tests displayed that the recognition precision was not considerably deteriorated by the hidden information. In addition, the losses of the enactment of the used biometric authentication mechanism were fewer for female than for male individuals.

2) INTRODUCTION OF THE AUDIO WATERMARKING SCHEME FOR DEEP LEARNING BASED AUDIO CONTENT CHARACTERIZATION APPLICATIONS

The sound signals that can be encountered in an indexing application are highly diversified according to the nature of the document to be processed, ranging from music to speech. Therefore, if speech analysis can be based on phonemes lasting few tens of milliseconds, music genre cannot be accurately perceived and classified only through a longer duration. Though automatic classification in sound classes has been deeply studied by the research community, most of the techniques proposed up to now only take into account partially the mechanism of human perception. Consequently, if their performance is acceptable for a particular classification problem, they are totally unacceptable for the other problems. Based on the hypothesis that humans are the best generalist classifiers of audio signal, the proposed approach suggest to inspire from the human perception mechanisms in order to develop automatic audio classification systems. We proposed a model of a hearing memory and a feature set of psychoacoustic inspiration.

Motivated by the great development of deep learning at the expense of the classic learning algorithms, we propose to combine the feature vector with the Deep Neural Networks to develop an audio classification system. Retrieved information characterizing the audio content is then embedded using an audio watermarking technique.

The proposed system is detailed and the adopted watermark embedding technique is also introduced. Experimental results are reported exhibiting the retrieved performance on public datasets. Finally, watermark robustness and transparency are assessed.

Remember that in the non-security watermarking applications, the robustness to intentional attacks is not required, a definite amount of robustness against licit treatments as the compression is necessary. In such applications, the watermark should commonly contain a great capacity information's and must be extracted using blind detection approach. One of the most popular applications for data transmission is sound document annotation. In fact, as we know, this application can be used to transfer a label to help signals indexing. The inserted information can contain meta-data describing the signal content or information about a target application. For example, the hidden watermark can indicate the name of the artist, the place of registration or any other data relating to the signal like in [52, 53].

In our case, the proposed watermarking technique DCT-MLP-LSB serves also to characterize the host audio document. A deep learning based strategy is exploited to analyze and classify the audio content into audio classes: music, speech, male speaker, happy speaker, etc. So a watermark containing information characterizing the audio content is constructed. In the extraction process, this watermark will inform about the audio class: music or speech, the speaker gender, etc.

3) PROPOSED SCHEME FOR CONTENT CHARACTERIZATION USING AUDIO WATERMARKING

Figure 13 summarize our scheme for content characterization based on deep learning using audio watermarking scheme. Main parts of the system are detailed: feature extraction, deep neural network classification and the watermarking scheme.

- Audio Feature extraction

Features must be the more informative conferring to the considered application. The audio file is usually divided into overlapping windows. Next, descriptors are calculated for each frame. Statics are made later in longer-term windows. So, we can define two processing levels, as displayed in the Figure 14: short-term and mid-term levels. Feature extraction that is a crucial stage in machine learning and pattern recognition tasks aims to envisage a set of features extracted from the considered dataset. As it is hard to directly perform on the original signal, feature extraction could be viewed as a data amount reduction procedure. In order to get a higher accuracy, it is imperative to select the most appropriate features set to the specific application.

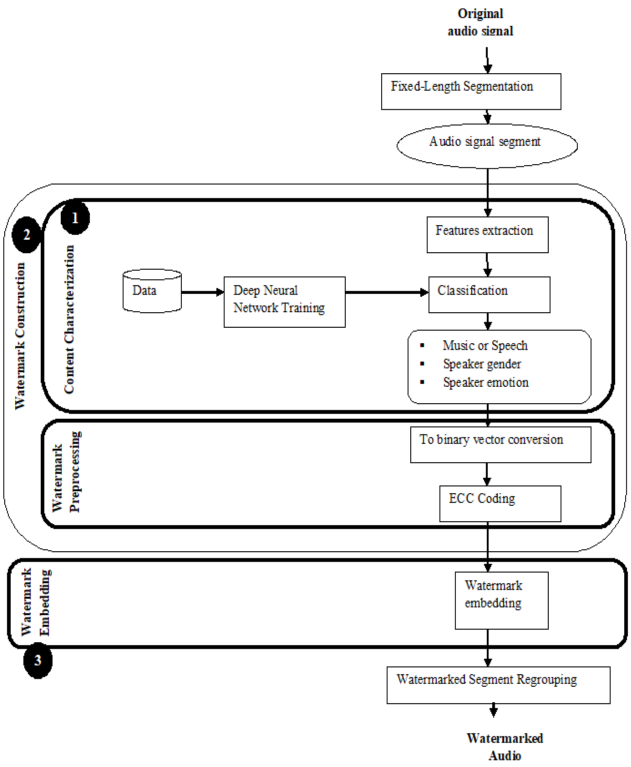


FIGURE 13. Proposed watermarking scheme for content characterization application

-Short-term analysis

In the short-term analysis known as frame-based processing, the audio file is divided into overlapping frames as exhibited in the Figure 14. The window duration at this level, is about 10 to 50 ms, within which the signal is considered as stationary. Consequently, descriptors can be extracted and computed during it [97]. After the framing step, a windowing is applied usually on each frame to evade discontinuities at block boundaries. In our approach, Hamming window is selected at this step. After windowing, the deliberated features will be calculated per frame as presented in the Figure 14. As stated by the computational way, extracted descriptors can be classified into time-domain features and frequency-domain ones.

Temporal Audio Features: these features are directly calculated from the audio samples. The most known time-domain features are Short-term energy [98] [99], energy entropy and zero crossing rate [98]. These features will be utilized in the feature extraction stage of our technique because they guarantee simple and good mean for audio signals analysis.

Spectral Audio Features: In order to guarantee correct audio analysis, it is essential to combine time-domain and frequency-domain features, called also spectral features. These metrics are computed using Discrete Fourier Transform (DFT) coefficients of the designed audio frame. The most known spectral-domain features are spectral flux , spectral centroid,[100], spectral roll off, Mel-Frequency Cepstrum Coefficients (MFCCs) [98,100], chroma vector [101] and Relative Spectral Analysis-Perceptual Linear Prediction (Rasta PLP) [102].

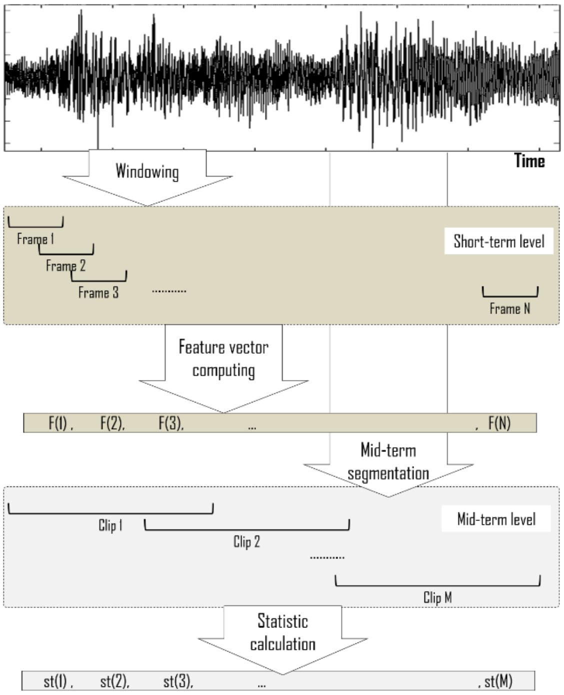


FIGURE 14. **Audio signal decomposition**

-Mid-term analysis

After performing the short-term known as the frame-based analysis, mid-term level statistics are computed. In effect, the frame-based processing was principally adopted in speech analysis as it was demonstrated that is more appropriate. Late, it was revealed that statistics made on longer-term windows could assure the semantic signification of the audio signal. Clip level or mid-term analysis is reached on probably overlapping fixed length segment fixed between 1 and 10 seconds. Clips represent a set of successive frames and depict the short-term features behavior. Indeed, the audio signal is separated into clips, and for each clip, statistics are calculated on the extracted short-term feature vector as illustrated in the Figure 14.

In this paper, we consider four mid-term statistics: the mean value, the standard deviation, the skewness and the kurtosis [103]. At first, each statistic metric is computed alone. After that, a fusion at feature level is proposed and performed between the proposed statistics as shown in Figure 15.

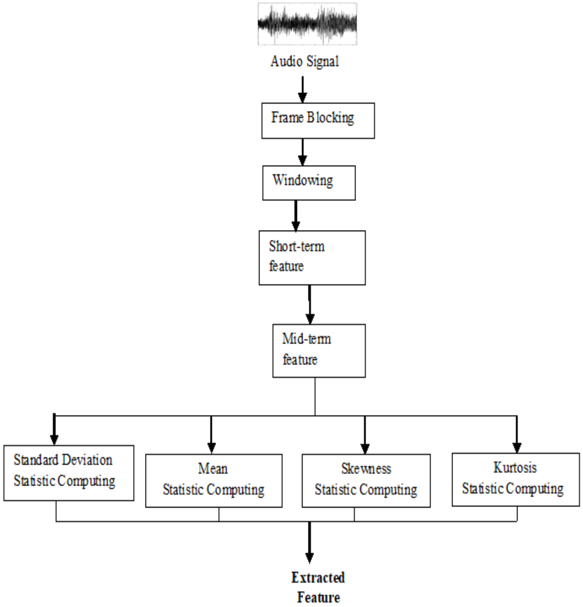


FIGURE 15. **Fusion at feature level**

- Deep learning based audio classification

In this work, we are interested in Deep Neural Networks DNNs which are interpretable deep neural networks such as a multilayer perceptron MLP as displayed in Figure 16. Block (1) of the Figure 17 targeting content characterization is performed using an MLP based architecture for audio classification for the three classification tasks: music and speech discrimination, speaker gender recognition and speech emotion identification. Categorical cross entropy is used as loss function and Softmax is used as activation function for the last dense layer.

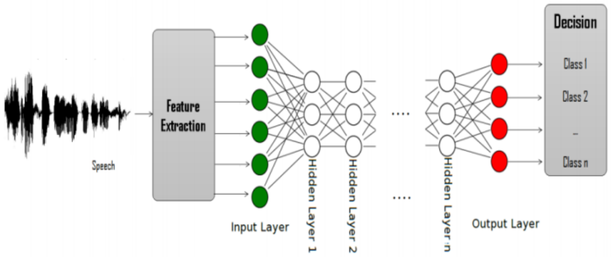


FIGURE 16. **MLP Deep Neural Network based audio classification scheme**

- Adopted technique of audio watermarking for deep learning based audio content characterization applications.

The proposed watermarking technique DCT-MLP-LSB as illustrated in the Figure 17 serves to characterize the host audio document. Indeed, at each segment, detected watermark will inform about the audio class: music or speech, the speaker gender, etc. We start by detailing the watermark construction block, and then we move to the mark hiding process [5]. The original file is first split into fixed length segment. Each segment is analyzed and classified into audio classes: music, speech, male speaker, happy speaker, etc. Then, the retrieved information characterizing the audio content will be inserted in the same signal. A binary vector is constructed using this information as following: for example, 0 is assigned to music and 1 to speech, etc. After that, and in order to improve the robustness propriety, a Hamming encoder (8, 12) is applied. Simultaneously, audio signal is divided into fixed length blocks of 512 samples. Each block is transformed in the spectral domain by using DCT. Embedding region is selected in the middle frequency band. Mean DCT value of this band is computed. The nearest frequency to this mean value is elected as the position of insertion. The Least Significant Bit LSB of this position is then replaced by the watermark bit value. Inverse DCT is after that applied. This process is repeated for each block along the audio stream. Thereby, each watermarked audio segment holds information about its content: music, speech, male speaker, speech emotion, etc. Detecting the watermark will allow to get these data and point to a moment according to a given criterion.

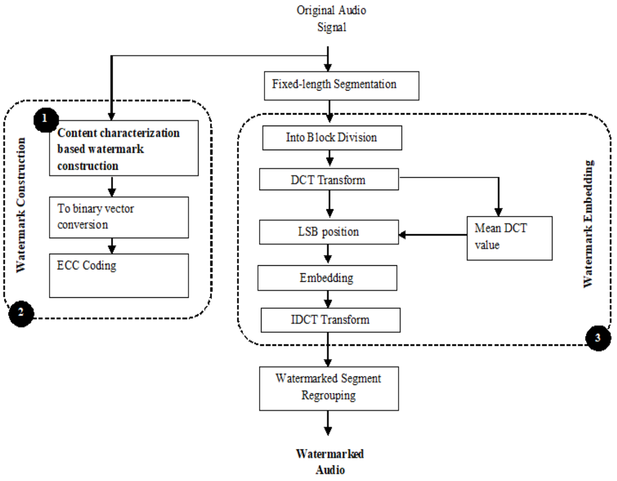


FIGURE 17. Proposed watermarking scheme of audio watermarking for deep learning based audio content characterization applications

3.2.4 Experimental results of the non-security watermarking application

- Classification assessment

In the next subsections, experimental results are reported for each task, using public datasets.

- *Experiments on speech music discrimination*

Two popular public databases were experimented: GTZAN and S&S Music/Speech datasets. The GTZAN corpus is collection of speech tracks and music excerpts assembled for classification purposes [104]. It involves 128 extracts lasting 30 seconds. They are mono 16-bit audio wav files sampled at 22050Hz. This dataset comprises various musical styles and speech tracks recorded in different conditions. The second corpus is Scheirer-Slaney (S&S) Music/Speech dataset [105]. It consists of a collection of 246 audio files saved in WAVE format and during 15 seconds each one. These extracts were collected at random from the radio including music and speech. Experimental results on the two databases are reported in table 11. Best achievement for the two datasets is attained by fusion between all statistics and using 50 neurons. Standard deviation outperforms other statics when undertaken without fusion in all cases for the two datasets. Achieved performance of prior approaches are depicted in table 12. We notice that the suggested scheme attains the higher performance.

- *Experiments on speaker gender identification*

The proposed system was experimented using two datasets in different languages: Eustace in English [106] and Berlin in German [107]. According to the table 13, gathering all statistics allow to enhance classification accuracies. In fact, for the first dataset, best achievement is obtained in case of statistics fusion besides in case of computing one statistic standard deviation or mean values. Unlike the first database, highest performance for the second dataset is achieved when all mid-term level statistics are fused and using 10 neurons for the three hidden layer. Unlike the first task, mean value outperform the other single statistics of this task. Achieved performance of some previous works are reported in table 14. It could be confirmed according to this table that the proposed scheme outperforms sate of the art approaches and afford promising results.

- *Experiments on speaker emotion recognition*

Two public datasets are used : Berlin Database of Emotional Speech and Surrey Audio Visual Expressed Emotion (SAVEE) database. In order to evaluate system performance, accuracy for each affective state is reported in the table 15 and 16. Best rate is obtained in all cases of neurons number when using all mid-term level statistics and highest values are achieved in case of 100 neurons for both databases. In case of single statistic, highest performance is achieved in case of mean value for the first database while the standard deviation outperforms other statistics for the second dataset. Through table 17, we notice that the suggested technique achieves promising recognition rates compared to the state of the art.

TABLE 11
CLASSIFICATION ACCURACY RESULTS FOR MUSIC/SPEECH DISCRIMINATION

	Classifier	Statistics	Music	Speech	Global
GTZAN Dataset	Deep (10,10,10)	Standard deviation	95%	95%	95%
		Mean	95%	85%	90%
		Skewness	85%	90%	87.5%
		Kurtosis	80%	95%	87.5%
		All statistics	100%	90%	95%
	Deep (50,50,50)	Standard deviation	95%	95%	95%
		Mean	90%	80%	85%
		Skewness	90%	95%	92.5%
		Kurtosis	95%	95%	95%
		All statistics	100%	95%	97.5%
	Deep (100,100,100)	Standard deviation	95%	95%	95%
		Mean	90%	100%	95%
		Skewness	90%	95%	92.5%
		Kurtosis	90%	95%	92.5%
		All statistics	95%	95%	95%
S&S Dataset	Deep (10,10,10)	Standard deviation	85%	100%	92.5
		Mean	85%	100%	92.5%
		Skewness	85%	95%	90%
		Kurtosis	75%	90%	82.5%
		All statistics	90%	100%	95%
	Deep (50,50,50)	Standard deviation	100%	100%	100%
		Mean	85%	100%	92.5%
		Skewness	90%	100%	95%
		Kurtosis	95%	100%	97.5%
		All statistics	100%	100%	100%
	Deep (100,100,100)	Standard deviation	95%	100%	97.5%
		Mean	80%	100%	90%
		Skewness	85%	100%	92.5%
		Kurtosis	90%	100%	95%
		All statistics	95%	100%	97.5%

TABLE 12
COMPARISON OF ACCURACY RESULTS FOR MUSIC/SPEECH DISCRIMINATION WITH PREVIOUS WORK

References	Best Acc rate
[108]	96.75%
[100]	93.5%
[109]	94.5%
[110]	95.9%
[111]	97.22%
[112]	97.28%
Our work on GTZAN	100%
Our work on S&S	100%

TABLE 13
CLASSIFICATION ACCURACY RESULTS FOR SPEAKER GENDER IDENTIFICATION

	Classifier	Statistics	Female	Male	Global
Eustace Dataset	Deep (10,10,10)	Standard deviation	100%	100%	100%
		Mean	100%	100%	100%
		skewness	100%	93.8%	96.9%
		Kurtosis	100%	87.5%	93.8%
		All statistics	100%	100%	100%
	Deep (50,50,50)	Standard deviation	100%	100%	100%
		Mean	100%	100%	100%
		skewness	100%	93.8%	96.9%
		Kurtosis	100%	91.7%	95.8
		All statistics	100%	100%	100%
	Deep (100,100,100)	Standard deviation	100%	100%	100%
		Mean	100%	100%	100%
		Skewness	100%	93.8%	96.9%
		Kurtosis	100%	93.8%	96.9%
		All statistics	100%	100%	100%
Berlin Dataset	Deep (10,10,10)	Standard deviation	84.5%	75.9%	80.2%
		Mean	96.6%	98.3%	97.4%
		Skewness	91.4%	81%	86.2%
		Kurtosis	91.4%	75.9%	83.6%
		All statistics	100%	97.8%	98.9%
	Deep (50,50,50)	Standard deviation	93.1%	65.5%	79.3%
		Mean	96.6%	94.8%	95.7%
		Skewness	87.9%	84.5%	86.2%
		Kurtosis	87.9%	72.4%	80.2%
		All statistics	93.5%	97.8%	95.7%
	Deep (100,100,100)	Standard deviation	84.5%	74.1%	79.3%
		Mean	94.8%	94.8%	94.8%
		Skewness	87.9%	84.5%	86.2%
		Kurtosis	89.7%	79.3%	84.5%
		All statistics	95.7%	97.8%	96.7%

TABLE 14.
COMPARISON OF ACCURACY RESULTS FOR SPEAKER GENDER IDENTIFICATION WITH PREVIOUS WORK

Ref	**Best Acc Rate**
[113]	95%
[114]	98.65%
[115]	90.1%
Our Work-Eustace dataset	100%
Our work-Berlin dataset	98.9%

TABLE 15
CLASSIFICATION ACCURACY RESULTS FOR EMOTION SPEECH RECOGNITION ON BERLIN DATASET

Dataset	Classifier	Statistics	Fear	Disgust	Happiness	Boredom	Neutral	Sadness	Anger	Global
Berlin Dataset	Deep (10,10,10)	Standard Deviatio	66.7%	44.4%	33.3%	33.3%	11.1%	55.6%	44.4%	41.3%
		Mean	22.2%	44.4%	55.6%	22.2%	33.3%	88.9%	33.3%	42.9%
		Skewness	33.3%	44.4%	33.3%	44.4%	55.6%	55.6%	66.7%	47.6%
		Kurtosis	44.4%	22.2%	44.4%	33.3%	33.3%	55.6%	55.6%	41.3%
		All	55.6%	88.9%	11.1%	11.1%	33.3%	88.9%	66.7%	50.8%
	Deep (50,50,50)	Standard Deviatio	77.8%	22.2%	22.2%	11.1%	66.7%	77.8%	33.3%	44.4%
		Mean	44.4%	55.6%	55.6%	22.2%	22.2%	66.7%	66.7%	47.1%
		Skewness	33.3%	55.6%	55.6%	66.7%	11.1%	0%	44.4%	38.1%
		Kurtosis	44.4%	66.7%	11.1%	22.2%	11.1%	55.6%	33.3%	34.9%
		All	66.7%	77.8%	55.6%	22.2%	22.2%	77.8%	77.8%	57.1%
	Deep (100,100,100)	Standard deviation	66.7%	33.3%	44.4%	11.1%	33.3%	77.8%	55.6%	46%
		Mean	66.7%	66.7%	55.6%	55.6%	11.1%	77.8%	66.7%	57.1%
		Skewness	55.6%	55.6%	77.8%	66.7%	11.1%	0%	22.2%	41.3%
		Kurtosis	33.3%	44.4%	44.4%	44.4%	33.3%	44.4%	55.6%	42.9%
		All	44.4%	88.9%	77.8%	33.3%	22.2%	88.9%	77.8%	61.9%

TABLE 16
CLASSIFICATION ACCURACY RESULTS FOR EMOTION SPEECH RECOGNITION ON SAVEE DATASET

Dataset	Classifier	Statistics	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Global
SAVEE Dataset	Deep (10,10,10)	Standard deviation	41.7%	66.7%	41.7%	33.3%	75%	83.3%	56.9%
		Mean	41.7%	50%	58.3%	33.3%	41.7%	58.3%	47.2%
		skewness	66.7%	16.7%	0%	8.3%	58.3%	16.7%	27.8%
		Kurtosis	75%	8.3%	25%	8.3%	50%	0%	27.8%
		All	58.3%	41.7%	50%	83.3%	58.3%	50%	56.9%
	Deep (50,50,50)	Standard deviation	50%	58.3%	50%	83.3%	100%	66.7%	68.1%
		Mean	75%	83.3%	41.7%	50%	50%	50%	58.3%
		skewness	66.7%	25%	8.3%	16.7%	41.7%	8.3%	27.8%
		Kurtosis	75%	8.3%	0%	8.3%	50%	16.7%	26.4%
		All	83.3%	33.3%	66.7%	58.3%	58.3%	75%	62.5%
	Deep (100,100,100)	Standard deviation	41.7%	50%	66.7%	58.3%	91.7%	66.7%	62.5%
		Mean	75%	58.3%	41.7%	50%	58.3%	66.7%	58.3%
		skewness	75%	8.3%	8.3%	16.7%	33.3%	8.3%	25%
		Kurtosis	66.7%	16.7%	16.7%	8.3%	50%	0%	26.4%
		All	83.3%	75%	66.7%	58.3%	66.7%	100%	75%

TABLE 17
COMPARISON OF ACCURACY RESULTS FOR EMOTION SPEECH RECOGNITION WITH PREVIOUS WORK

Ref	Best Acc Rate
[116]	49%
[117]	53%
[118]	72.05%
[119]	71.7%
[120]	76.3%
Our work-Berlin Dataset	61.9%
Our work-SAVEE Dataset	75%

- Watermarking evaluation

After audio analysis process assessment, watermarking algorithm is evaluated in term of transparency and robustness.

- *Watermarking transparency results*

Signal to noise ratio SNR, comparing between the original and the watermarked files, is computed. According to the recommendation of the IFPI, transparency is confirmed when SNR values exceeded 20dB. Reported results are presented in Figure 18. According to the achieved SNR, watermark transparency is confirmed by very high values greater than 40 in all cases.

- *Watermarking robustness results*

Since any signal is compressed before storage or transmission, watermarking scheme should resist to such transformation and watermark should be correctly detected even after compression. MP3 encoder is experimented using typical compression ratio since it is the most utilized audio encoder. As shown in Figure 19, NC values are higher than 0.8 confirming that the mark is almost detected. Robustness against StirMark attacks is then assessed as shown in Figure 20. NC values are equal to 1 in most cases confirming the robustness of the proposed watermarking algorithm to the majority of attacks; excepting the cases of Add noise 700 and Dynnoise attacks where the NC is slightly lower than 1. This problem could be circumvented by the mark duplication.

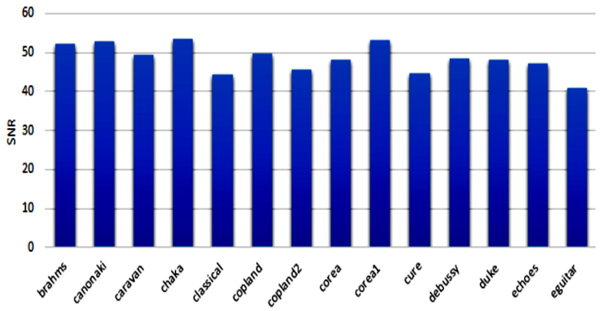


FIGURE 18. SNR values after audio watermarking for deep learning based audio content characterization applications

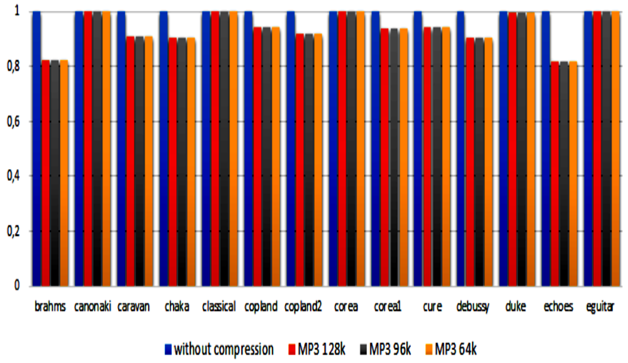


Figure 19. NC values after compression attacks in audio watermarking for deep learning based audio content characterization applications

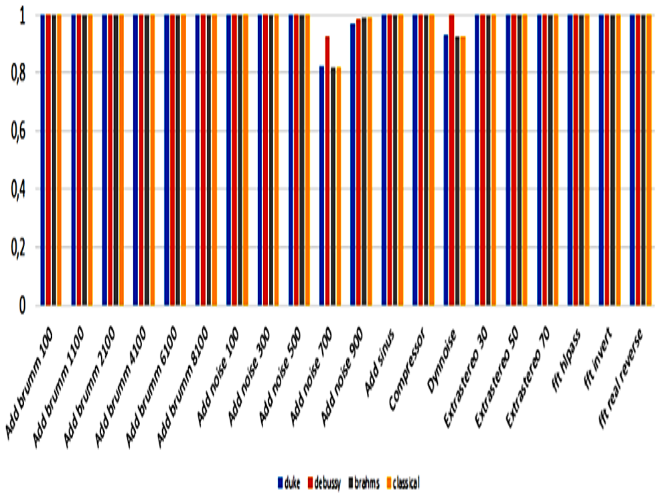


Figure 20. NC values after Stirmark attacks in audio watermarking for deep learning based audio content characterization applications

IV. CONCLUSION

Digital audio watermarking can be used in different types of applications that target two different situations; the first one for security applications and the second one for non-security applications. Thus, in this paper, we carried a big attention in examining these situations. Then we proposed two digital watermarking schemes that we have implemented for basic and sensitive digital contents. A first scheme is an audio watermarking technique for security copyright protection application. This first work is hiding the signature in a narrow middle frequency band of an audio frame. We have involved NN architecture in the proposed insertion and detection processes to improve security and robustness even with high watermark capacity. Furthermore, we have studied and integrated some masking phenomena of the HPM. The objective is to determine the masking threshold curve and to compare it with the estimated Power Spectrum Density envelope to insert appropriately the signature under this curve. Experimental results have proved that using frequency perceptual masking with the spectral envelope estimation in the frequency domain offer good robustness results comparing with our previous NN based audio watermarking technique [2] and with other existing watermarking techniques. In summary, we can endorse that we have implemented an audio watermarking scheme meeting the requirements set by the IFPI with good robustness and imperceptibility results. Moreover, our proposed audio watermarking scheme is very useful for copyright protection of standard audio files and also sensitive audio data like Quranic files but can also be extended to guarantee content integrity verification, proof of authenticity and tamper detection of those signals.

Furthermore, we have suggested a second new audio watermarking approach for content characterization as non-security application. The originality consists in using watermark holding information characterization the audio content. Once detected, the user could browse the audio file and move to a selected moment according to given criteria. For example, speech segments, uttered by a male speaker with a happy emotional state, could be picked out. For audio content analysis and classification, a deep learning based scheme was adopted and combined to a rich descriptor set. Moreover, for watermarking, a frequency domain technique is employed based on DCT transform. Reported results showed that the proposed scheme achieves higher performance at classification level as well as at watermarking.

As we are very interested with new digital watermarking applications, we are focusing in adopting our proposed audio watermarking schemes for video content to propose solution combating fake data such as fake election news or fake covid-19 related news.

ACKNOWLEDGMENT

This research was supported by Taif University Researchers Supporting Project number (TURSP-2020/216), Taif University, Taif, Saudi Arabia for supporting this work.

REFERENCES

[1] Rizzo, S., Bertini, F. & Montesi, D. Fine-grain watermarking for intellectual property protection. EURASIP J. on Info. Security 2019, 10.

[2] Charfeddine M., El'arbi M., Koubaa M., Ben Amar C.. DCT based blind audio watermarking scheme, International conference on signal processing and multimedia applications SIGMAP, Athens-Greece, 2010, p. 139–144.

[3] Bhat V., Sen Gupta I., Das A.. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain , Digital Signal Processing, 2010, vol 20, no 6, p. 426–436.

[4] Charfeddine M., El'Arbi M., Ben Amar C.. A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain, In Proceeding of the 3rd international symposium on communications, control and signal processing ISCCSP'03, Malta, 2008, p. 1138–1143.

[5] Charfeddine M., Masmoudi S., Bellaaj M., Ben Amar C.. Un schéma aveugle de tatouage audio numérique opérant sur les bits les moins significatifs dans le domaine fréquentiel utilisant un code correcteur d'erreurs, 6èmes Ateliers de Traitement et Analyse de l'Information : Méthodes et Applications TAIMA'06, Hammamet-Tunisie, 2009, p. 371-377.

[6] Terchi, Y., Bouguezel, S. A blind audio watermarking technique based on a parametric quantization index modulation. MultimedTools Appl, 2018,77, 25681–25708.

[7] Dappuri, B., Rao, M.P. & Sikha, M.B. Non-blind RGB watermarking approach using SVD in translation invariant wavelet space with enhanced Grey-wolf optimizer. Multimed Tools Appl,2020.

[8] Masmoudi,S., Charfeddine, M. & Ben Amar,C., A robust audio watermarking technique based on the perceptual evaluation of audio quality algorithm in the multiresolution domain, The 10th IEEE International Symposium on Signal Processing and Information Technology, (ISSPIT 2010) Luxor, pp. 326-331, https://doi.org/10.1109/ISSPIT.2010.5711803

[9] T Y Liu, W H Tsai. Generic lossless visible watermarking - a new approach. IEEE Trans on Image Processing. 2010. 19(5). 1224-1235

[10] Qi, W., Yang, G., Zhang, T. et al. Improved reversible visible image watermarking based on HVS and ROI-selection. Multimed Tools Appl, 2019, 78, 8289–8310

[11] Tarhouni, N., Charfeddine, M. & Ben Amar, C. Novel and Robust Image Watermarking for Copyright Protection and Integrity Control. Circuits Syst Signal Process , 2020.

[12] Lancini R., Mapelli F., Tubaro S., A robust video watermarking technique for compression and transcoding processing , International Conference on Multimedia and Expo ICME, Lausanne, Switzerland,2002, p. 549-552.

[13] Quan X., Zhang H., Perceptual Criterion Based Fragile Audio Watermarking Using Adaptive Wavelet Packets , in proceedings of the Pattern Recognition ICPR, Cambridge, UK, 2004, vol 2, p. 867-870.

[14] M. Zamani and A. B. A. Manaf, Genetic algorithm for fragile audio watermarking, Telecommunication Systems, 2015, vol. 59, no. 3, pp. 291–304.

[15] Rakhmawati, L., Wirawan, W. & Suwadi, S. A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability. J Image Video Proc. 2019, 61

[16] M. Q. Fan, P. P. Liu, H. X. Wang, and H. J. Li, A semi-fragile watermarking scheme for authenticating audio signal based on dual-tree complex wavelet transform and discrete cosine transform, International Journal of Computer Mathematics, 2013, vol. 90, no. 12, pp. 2588-2602.

[17] S. Zhaopin et al., Window switching strategy based semi-fragile watermarking for MP3 tamper detection. Multimed. Tools Appl. 2017, 76(7), 9363–9386.

[18] Taranovsky D. Data Hiding and Digital Watermarking. In: Chen J., Cranton W., Fihn M. (eds) Handbook of Visual Display Technology. Springer, Berlin, Heidelberg, 2012.

[19] Brar A.S., Kaur M. A Survey of Reversible Watermarking Techniques for Data Hiding with ROI-Tamper Detection in Medical Images. In: Das V.V., Chaba Y. (eds) Mobile Communication and Power Engineering. AIM 2012. Communications in Computer and Information Science,Springer, Berlin, Heidelberg, 2013, vol 296.

[20] Y. Xiong and Z. X. ming, Covert Communication Audio Watermarking Algorithm Based on LSB, International Conference on Communication Technology, Guilin, 2006, pp. 1-4.

[21] Li, F., Li, B., Huang, Y. et al. Research on covert communication channel based on modulation of common compressed speech codec. Neural Comput & Applic , 2020.

[22] Deshpande and J. Gadge, New Watermarking Technique for Relational Databases, Second International Conference on Emerging Trends in Engineering & Technology, Nagpur, 2009, pp. 664-669.

[23] Khanam, T.; Dhar, P.K.; Kowsar, S.; Kim, J.-M. SVD-Based Image Watermarking Using the Fast Walsh-Hadamard Transform, Key Mapping, and Coefficient Ordering for Ownership Protection. Symmetry 2020, 12, 52.

[24] L. Priya C.V. and N. Raj N.R., Digital watermarking scheme for image authentication, International Conference on Communication and Signal Processing (ICCSP), Chennai, 2017, pp. 2026-2030,

[25] Anand, A., Singh, A.K. Watermarking techniques for medical data authentication: a survey. Multimed Tools Appl , 2020.

[26] Chen, Y., Huang, H. Coevolutionary genetic watermarking for owner identification. Neural Comput & Applic, 2015, 26, 291–298.

[27] P. de Jesus Vega-Hernandez, M. Cedillo-Hernandez, M. Nakano, A. Cedillo-Hernandez and H. M. Perez-Meana, Ownership Identification of Digital Video via Unseen-Visible Watermarking, 7th International Workshop on Biometrics and Forensics (IWBF), Cancun, Mexico, 2019, pp. 1-6.

[28] S. Samuel and W. T. Penzhorn, Digital watermarking for copyright protection, IEEE Africon. 7th Africon Conference in Africa (IEEE Cat. No.04CH37590), Gaborone, 2004, pp. 953-957.

[29] F. Ernawan and M. N. Kabir, An Improved Watermarking Technique for Copyright Protection Based on Tchebichef Moments, in IEEE Access, 2019, vol. 7, pp. 151985-152003.

[30] Charfeddine, M., El'arbi, M. & Ben Amar, C. A new DCT audio watermarking scheme based on preliminary MP3 study. Multimed Tools Appl, 2014, 70, 1521–1557.

[31] El'Arbi, M., Koubaa, M., Charfeddine, M. et al. A dynamic video watermarking algorithm in fast motion areas in the wavelet domain. Multimed Tools Appl , 2011, 55, 579–600.

[32] N.Tarhouni, M.Charfddine, C.B.Amar, A New Robust and Blind Image Watermarking Scheme In Frequency Domain Based On Optimal Blocks Selection,International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 2018, 78-86.

[33] Zhaofeng, M., Weihua, H. & Hongmin, G. A new blockchain-based trusted DRM scheme for built-in content protection. J Image Video Proc. 2018, 91.

[34] Chaabane F., Charfeddine, M., Puech ,W., Ben Amar ,C., A survey on digital tracing traitors schemes, 9th International Conference on Information Assurance and Security (IAS 2013), Gammarth, pp. 85-90, https://doi.org/10.1109/ISIAS.2013.6947738

[35] Franco-Contreras J., Coatrieux G. Databases Traceability by Means of Watermarking with Optimized Detection. In: Shi Y., Kim H., Perez-Gonzalez F., Liu F. (eds) Digital Forensics and Watermarking. IWDW,Lecture Notes in Computer Science, vol 10082. Springer, Cham. 2017.

[36] Chaabane F., Charfeddine M., Puech W., Amar C.B. Towards a Blind MAP-Based Traitor Tracing Scheme for Hierarchical Fingerprints. In: Arik S., Huang T., Lai W., Liu Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science, vol 9492. Springer, Cham. 2015.

[37] Chaabane, F., Charfeddine, M., Puech, W. et al. A two-stage traitor tracing scheme for hierarchical fingerprints. Multimed Tools Appl 76, 14405–14435 ,2017.

[38] Chaabane, F., Charfeddine, M., Puech,W., Ben Amar, C., A QR-code based audio watermarking technique for tracing traitors, 23rd European Signal Processing Conference (EUSIPCO 2015), Nice, pp. 51-55, https://doi.org/10.1109/EUSIPCO.2015.7362343

[39] Zhang, Guoyin & Kou, Liang & Zhang, Liguo & Liu, Chao & Da, Qingan & Sun, Jianguo. A New Digital Watermarking Method for Data Integrity Protection in the Perception Layer of IoT. Security and Communication Networks. 2017.

[40] V. Choudhary, M. K. Dutta and A. Singh, Reversible Watermarking scheme for Authentication and integrity control in Biometric Images, 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2019, pp. 662-666.

[41] Masmoudi, S., Charfeddine, M. & Ben Amar, C. A Semi-Fragile Digital Audio Watermarking Scheme for MP3-Encoded Signals Using Huffman Data. Circuits Syst Signal Process, 2020, 39, 3019–3034.

[42] H.Bay, A.ESS, T.TUYTELAARS, L.GOOL, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), 2008, 346-359.

[43] J. A. Bloom, I. J. Cox, T. Kalker, J. -. M. G. Linnartz, M. L. Miller and C. B. S. Traw, Copy protection for DVD video, in Proceedings of the IEEE, vol. 87, no. 7, 1999, pp. 1267-1276.

[44] M. Maes, T. Kalker, J. -. M. G. Linnartz, J. Talstra, F. G. Depovere and J. Haitsma, Digital watermarking for DVD video copy protection, in IEEE Signal Processing Magazine,2000, vol. 17, no. 5, pp. 47-57.

[45] R. Petrovic and V. Atti, Watermark based access control to copyrighted content, 2013 11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Nis, 2013, pp. 315-322.

[46] Abd-El-Atty B, Iliyasu AM, Alaskar H, Abd El-Latif AA. A Robust Quasi-Quantum Walks-Based Steganography Protocol for Secure Transmission of Images on Cloud-Based E-healthcare Platforms. Sensors (Basel). 2020;20(11):3108.

[47] G. Zhou and D. Lv, An Overview of Digital Watermarking in Image Forensics, 2011 Fourth International Joint Conference on Computational Sciences and Optimization, Yunnan, 2011, pp. 332-335.

[48] Qasim, A.F., Aspin, R., Meziane, F. et al. ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images. Multimed Tools Appl, 2019, 78, 16433–16463

[49] L. Liu and X. Li, Watermarking Protocol for Broadcast Monitoring, 2010 International Conference on E-Business and E-Government, Guangzhou, 2010, pp. 1634-1637.

[50] M. Parvaix, L. Girin and J. Brossier, A Watermarking-Based Method for Informed Source Separation of Audio Signals With a Single Sensor, in IEEE Transactions on Audio, Speech, and Language Processing, 2010, vol. 18, no. 6, pp. 1464-1475.

[51] Bailly, G., Attina, V., Baras, C., Bas, P., Baudry, S., Beautemps, D., Brun, R., Chassery, J.-M., Davoine, F., Elisei, F., Gibert, G., Girin, L., Grison, D., Léoni, J.-P., Liénard, J., Moreau, N. & Nguyen, P. ARTUS: synthesis and audiovisual watermarking of the movements of a virtual agent interpreting subtitling using Cued Speech for deaf televiewersAMSE - Advances in Modelling, 2007, 67, pp 177-187.

[52] Tzanetakis G, Music information retrieval: theory and applications. In: Proceedings of the 17th ACM international conference on Multimedia, ACM, 2009, pp 915–916

[53] Mezghani E, Charfeddine M, Nicolas H, Ben Amar C (2015) Audiovisual video characterization using audio watermarking scheme. In: ISDA, 2015 15th International Conference on Intelligent Systems Design and Applications, IEEE, 2015, pp 213–218

[54] Hirakawa, Manabu & Iijima, J.. Mobile Services and Implementation of Digital Watermarks in Audio Files. IIHMSP.2010, 31, 94-97. 10.1109

[55] Agbaje, M. O., Awodele, O., & Ogbonna, A. C. Big data, audience measurement and digital watermarking: A review. Proceedings of the e-Skills for Knowledge Production and Innovation Conference, Cape Town, South Africa, 2014, 17-28.

[56] Portilla, I., Television Audience Measurement: Proposals of the Industry in the Era of Digitalization, Trípodos, 2015, N.36, pp. 75-92

[57] Prior F, Ingeholm ML, Levine BA, Tarbox L: Potential impact of HITECH security regulations on medical imaging. In: 2009 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC 2009, Piscataway, NJ, USA, 2009, pp. 2157–60

[58] Ruotsalainen P: Privacy and security in teleradiology. European Journal of Radiology, 2010, 73:31–35.

[59] Yazeed Alkhurayyif, National ID Cards, International Journal of Computing Science and Information Technology, 2013Vol.1 (02) 44 – 48, ISSN: 2278-9669.

[60] S. Hakak, A. Kamsin, O. Tayan, M. Y. Idna Idris, A. Gani and S. Zerdoumi, Preserving Content Integrity of Digital Holy Quran: Survey and Open Challenges, in IEEE Access, 2017, vol. 5, pp. 7305-7325.

[61] F. Kurniawan, M. S. Khalil, M. K. Khan and Y. M. Alginahi, Exploiting Digital Watermarking to Preserve Integrity of the Digital Holy Quran Images, 2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences, Madinah, 2013, pp. 30-36.

[62] Kamaruddin, Nurul & Kamsin, Amirrudin & Hakak, Saqib. (2017). Associated diacritical watermarking approach to protect sensitive arabic digital texts. AIP Conference Proceedings. 2017, 1891.

[63] H. Hu and T. Lee, "Hybrid Blind Audio Watermarking for Proprietary Protection, Tamper Proofing, and Self-Recovery," in *IEEE Access*, vol. 7, pp. 180395-180408, 2019, doi: 10.1109/ACCESS.2019.2958095.

[64] Wu, Q.L.; Wu, M. A Novel Robust Audio Watermarking Algorithm by Modifying the Average Amplitude in Transform Domain. Appl. Sci. 2018, 8, 723.

[65] Lanxun W., Chao Y., Jiao P., An audio watermark embedding algorithm based on meanquantization in wavelet domain , In Proceedings of the 8th international conference on electronic measurement and instruments, Xi'an, China, 2007, vol. 2, p. 423–425

[66] Wu, Q.; Wu, M. Adaptive and Blind Audio Watermarking Algorithm Based on Chaotic Encryption in Hybrid Domain. Symmetry 2018, 10, 284. https://doi.org/10.3390/sym10070284

[67] H.-T. Hu and L.-Y. Hsu, A DWT-based rational dither modulation scheme for effective blind audio watermarking, Circuits, Systems, and Signal Processing, 2016 vol. 35, no. 2, pp. 553–572.

[68] Jinquan Zhang, Bin Han: Robust Audio Watermarking Algorithm Based on Moving Average and DCT. CoRRabs, 2017, 1704.02755

[69] Pan D. A tutorial on MPEG/audio compression, in IEEE MultiMedia, vol. 2, no. 2, pp. 60 - 74, Summer 1995, doi: 10. 1109/93. 388209.

[70] Markel J.D. and Gray A.H., Linear Prediction of Speech, Journal of Sound and Vibration, Springer, 1976, vol 51, no 4, p. 595-595.

[71] Dittmann, J., Kraetzer, C. (eds.): ECRYPT Deliverable D.WVL.10 - Audio Benchmarking Tools and Steganalysis; Rev. 1.1, 2006.

[72] Kutter M., Petitcolas F.A.P., A fair benchmark for image watermarking systems, the electronic Imaging'99 - Security and Watermarking of Multimedia Contents, San Jose, CA, 1999, vol 3657, p. 226-239.

[73] Kutter M., Hartung F., Introduction to watermarking techniques, In F.A.P. Petitcolas & S. Katzenbeisser (Eds.), Information hiding: Techniques for steganographie and digital watermarking (1 ed.) Artech House, Boston,2000, p. 97-120.

[74] Gan W.-S., Kuo S.-M., Embedded signal processing with the Micro Signal Architecture, Wiley-IEEE Press., ISBN 978-0-471-73841-1, 2007.

[75] Xiaojuan X., Peng H., He C., DWT-based audio watermarking using support vector regression and subsampling , In Proceedings of the 7th international workshop on fuzzy logic and applications, 2007, p. 136–144.

[76] El'Arbi, M., Charfeddine, M., Masmoudi, S., Koubaa, M. & Ben Amar, C., Video watermarking algorithm with BCH error correcting codes hidden in audio channel, in Proceedings of the, IEEE Symposium Series in Computational Intelligence. Paris (SSCI 2011), pp. 164–17

[77] H.-T. Hu, L.-Y. Hsu, H.-H. Chou, Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity. 2008, Digit. Signal Process. 31, 115–123

[78] Lu, W.; Chen, Z.; Li, L.; Cao, X.; Wei, J.; Xiong, N.; Li, J.; Dang, J. Watermarking Based on Compressive Sensing for Digital Speech Detection and Recovery †. Sensors 2018, 18, 2390. https://doi.org/10.3390/s18072390

[79] Christian N., Jurgen H., Digital watermarking and its influence on audio quality, In Proceeding of the 105th AES convention, 1998.

[80] Acevedo A. G., Audio Watermarking Quality Evaluation, e - Business and Telecommunication Networks (Edited by J. Ascenso et al.), Springer, 2006, Netherlands, p. 272 - 283.

[81] H.-T. Hu, L.-Y. Hsu, H.-H. Chou, Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity. 2008, Digit. Signal Process. 31, 115–123

[82] Bhat V., Sen Gupta I. and Das A., Audio watermarking based on quantization in wavelet domain, In Proceedings ICISS'08, LNCS-5352, 2008, p. 235-242.

[83] Martinez-Noriega R., Kang H., Kurkoski B., Yamaguchi K., Kobayashi K., Nakano M., Increasing robustness of audio watermarking DM using ATHC codes, In Proceedings of the Mexican conference on informatics security, Oaxaca, Mexico 2006.

[84] Ali A.-H., Ahmad M., Digital audio watermarking based on the discrete wavelets transform and singular value decomposition , Eur J Sci Res 39(1), 2010, p. 6–21

[85] Sehirli M., Gürgen F.S., Ikizoglu S. , Performance evaluation of digital audio watermarking techniques designed in time, frequency and cepstrum domains , In Proceeding of the international conference on advances in information systems,Izmir, Turkey, 2004, vol 3261, p. 430–440.

[86] Uludag U., Arslan L.-M., Audio watermarking using DC level shifting, Advanced Topics in Speech Processing Project Report, Electrical and Electronics Engineering Department. Bogazici University, Turkey, 2001, p. 1–6.

[87] Bender W., Gruhl D., Morimoto N. and Lu, A., Techniques for data hiding, IBM Systems Journal, 1996, vol. 35, no. 3-4, p. 313-336.

[88] Swanson M., Zhu B., Tewfic A., Boney L., Robust audio watermarking using perceptual masking, Signal Processing, 2008, vol 66, no 3, p. 337–355.

[89] Kais Khaldi, Abdel-Ouahab Boudraa, Audio Watermarking Via EMD. IEEE Trans. Audio, Speech & Language Processing,2013, 21(3) 675-680(2013)

[90] Pranab Kumar Dhar and Tetsuya Shimamura ,A Blind LWT-Based Audio Watermarking Using Fast Walsh Hadamard Transform and Singular Value Decomposition, International Symposium on Circuits and Systems, IEEE, 2014.

[91] Aree A. Mohammed, Robust Audio Watermarking Comparison between Modified Phase and Wavelet Based Techniques,International Journal of Computer Science Engineering,2015, Vol. 4, No. 6.

[92] W. Diao, Y. Wu, W. Zhang, B. Liu and N. Yu, Robust Audio Watermarking Algorithm Based on Air Channel Characteristics, 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, 2018, pp. 288-293.

[93] Esmaili, Shahrzad (2004): Content Based Audio Watermarking and Retrieval Using Time-Frequency Analysis. Ryerson University. Thesis. https://doi.org/10.32920/ryerson.14655894.v1

[94] Umapathy, Karthikeyan & Ghoraani, Behnaz & Krishnan, Sridhar. (2010). Audio Signal Processing Using Time-Frequency Approaches: Coding, Classification, Fingerprinting, and Watermarking. Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing. 28. 10.1155/2010/451695.

[95] Nematollahi, Mohammad Ali & Al-Haddad, Syed Abdul Rahman & Doraisamy, Shyamala & Gamboa-Rosales, Hamurabi. (2016). Speaker Frame Selection for Digital Speech Watermarking. National Academy Science Letters. 39. 10.1007/s40009-016-0430-8.

[96] Andrea Oermann, Andreas Lang, Claus Vielhauer; "Digital Speech Watermarking and its Impact to Biometric Speech Authentication", In Jana Dittmann Claus Vielhauer, Jan Hansen, New Advances in Multimedia Security, Biometrics, Watermarking and Cultural Aspects, Logos Verlag Berlin 2006, pp. 33-51.

[97] Wang Y, Liu Z, Huang JC Multimedia content analysis-using both audio and visual clues. Signal Processing Magazine, 2000, IEEE 17(6):12–36

[98] Vavrek J, Voz´arikov´a E, Pleva M, Juha´r J (2012) Broadcast news audio classification using svm binary trees. In: Telecommunications and Signal Processing (TSP), 35th International Conference on, IEEE, 2012, pp 469–473

[99] Mezghani, E, Charfeddine , M., & Ben Amar, C., Audio silence deletion before and after MPEG video compression, International Conference on Computer Applications Technology (ICCAT 2013), Sousse, pp. 1-5, https://doi.org/10.1109/ICCAT.2013.6521969

[100] Sell G, Clark P (2014) Music tonality features for speech/music discrimination. In: Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on, 2014, IEEE, pp 2489–2493

[101] Becker J, Rohlfing C A segmental spectral flatness measure for harmonic-percussive discrimination, 2013.

[102] Hermansky H, Morgan N, Bayya A, Kohn P (1992) Rasta-plp speech analysis technique. In: Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, IEEE, vol 1, pp 121–124

[103] Baniya BK, Lee J Importance of audio feature reduction in automatic music genre classification. Multimedia Tools and Applications2016, 75(6):3013–302

[104] Tzanetakis G, Cook P Sound analysis using mpeg compressed audio. In: ICASSP'00. Proceedings. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, 2000, vol 2, pp II761–II76

[105] Scheirer E, Slaney M (1997) Construction and evaluation of a robust multifeature speech/music discriminator. In: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on, IEEE, 1997, vol 2, pp 1331–1334

[106] White L, King S The eustace speech corpus (http://www.cstr.ed.ac.uk/projects/eustace). In: Centre for Speech Technology Research, University of Edinburgh, 2003.

[107] Burkhardt F, Paeschke A, Rolfes M, Sendlmeier WF, Weiss B A database of german emotional speech. In: Interspeech, 2005, vol 5, pp 1517–1520

[108] Khonglah BK, Prasanna SM Speech/music classification using speech-specific features. Digital Signal Processing, 2016, 48:71–83

[109] Panagiotakis C, Tziritas G A speech/music discriminator based on rms and zerocrossings., IEEE Transactionson Multimedia, 2005, 7(1):155–166

[110] El-Maleh K, Klein M, Petrucci G, Kabal P Speech/music discrimination for multimedia applications. In: Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on, 2000, IEEE, vol 6, pp 2445–2448

[111] Fu Zh, Wang JF Robust features for effective speech and music discrimination. In: ROCLING, 2008.

[112] Munoz-Exposito JE, Garcia-Galan S, Ruiz-Reyes N, Vera-Candeas P, Rivas-Pen˜a F Speech/music discrimination using a single warped lpc-based feature. In: Proc. ISMIR,, 2005, vol 5, pp 16–25

[113] Gaikwad S, Gawali B, Mehrotra S Gender identification using svm with combination of mfcc. Advances in Computational Research 4(1), 2012.

[114] Hu Y, Wu D, Nucci A Pitch-based gender identification with two-stage classification. Security and Communication Networks,2012, 5(2):211–225.

[115] Ahmad J, Fiaz M, Kwon Si, Sodanil M, Vo B, Baik SW Gender identification using mfcc for telephone applications-a comparative study. arXiv preprint arXiv:2016, 160101577

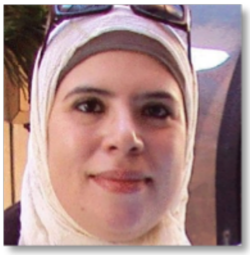
[116] Dai K, Fell HJ, MacAuslan JRecognizing emotion in speech using neural networks. Telehealth and Assistive Technologies, 2008, 31:38–43.

[117] Haq S, Jackson PJ, Edge J Audio-visual feature selection and reduction for emotion classification. In: Proc. Int. Conf. on Auditory-Visual Speech Processing (AVSP08), Tangalooma, Australia, 2008.

[118] Shah F, et al Discrete wavelet transforms and artificial neural networks for speech emotion recognition. International Journal of Computer Theory and Engineering, 2010, 2(3):319

[119] Javidi MM, Roshan EF Speech emotion recognition by using combinations of c5. 0, neural network (nn), and support vector machines (svm) classification methods. J Math Comput Sci,2013,6:191–200

[120] Sheikhan M, Bejani M, Gharavian D Modular neural-svm scheme for speech emotion recognition using anova feature selection method. Neural Computing and Applications, 2013, 23(1):215–22



MAHA CHARFEDDINE became a Member (M) of IEEE in 2008, a Senior She was born on 30th March 1981 and she has the Tunisian Nationality. She received her engineering diploma in computer sciences from the Tunisian university ENIS-SFAX (National School of Engineers ENIS, University of Sfax, Tunisia) in June 2005. In 2007, she obtained the Master and the Ph.D degrees in Computer Sciences, respectively in 2007 and in 2013, both from the ENIS-SFAX. Currently, Maha CHARFEDDINE is an Assistant Professor in the computer sciences and applied mathematics department at ENIS, University of Sfax. The topics of taught courses are mainly Cyber-security, Norms and standards of Multimedia Systems, IT project management and Linux Operating System. She directed many undergraduate projects of end studies. She is a member the Research Groups on Intelligent Machines of REGIM-Lab (LR11ES48). She co-supervises graduate students in Master and PhD degrees. Main research activities focus on Digital Watermarking for Copyright Protection, Traceability, Content Characterization, Integrity Control, Tamper Localization, Recovery and studies of Human Psychoacoustic/Visual Models of Audio and Image Standard coders and decoders. She participated and contributed, on 3 and 4 December 2020, to the fulfillment of the action plan of the National Cyber-Security Strategy 2020-2025.

EYA MEZGHANI was born in Ariana (Tunisia) in 1985. She obtained her Telecom Engineer Diploma in 2009 from the National Engineering School of Tunis (ENIT), Tunisia. She received the Master Diploma in computer sciences in 2012 from the National Engineering School of Sfax ENIS-SFAX. She obtained her Doctorate Degree in "Computer System Engineering" in 2018 from the ENIS-SFAX. She is a also research member in the Research Group of Intelligent Machine REGIM-Lab, (ENIS), Sfax (Tunisia). Her researches include audio analysis and classification, artificial intelligence, digital watermarking and signal processing.

MASMOUDI SALMA was born in Sfax(Tunisia) in 1984. She obtained her Engineering diploma in Computer science in 2008, and her master diploma in New Technology of the Dedicated Computer Sciences Systems in 2010, respectively all from the National School of Engineers of Sfax ENIS-SFAX. She has worked as contractual assistant in the National School of Engineers of Sfax from 2011 to 2014.She is currently working towards the PhD degree in Computing Systems Engineering in the laboratory REGIM-Lab in the National School of Engineers of Sfax, University of Sfax, Tunisia. Her research interests include Digital audio watermarking, MP3 Compression.

CHOKRI BEN AMAR received the B.S. degree in Electrical Engineering from the National Engineering School of Sfax (ENIS) in 1989, the M.S. and PhD degrees in Computer Engineering from the National Institute of Applied Sciences in Lyon, France, in 1990 and 1994, respectively. He spent one year at the University of "Haute Savoie" (France) as a teaching assistant and researcher before joining the higher School of Sciences and Techniques of Tunis (ESSTT) as Assistant Professor in 1995. In 1999, he joined the Sfax University (USS) as Assistant Professor, and since 2011 as a full professor in the Department of Computer Sciences and Applied Mathematics of the National Engineering School of Sfax. Since September 2018, he is a full professor at the college of Computers and Information technology of Taif University in Saudi Arabia.

His research interests include Computer Vision and Image and video analysis. These research activities are centered on intelligent algorithms and their applications to data Classification and approximation, Pattern Recognition, Watermarking and image and video indexing and securing. He is a senior member of IEEE since 2008. He founded the IEEE Signal Processing Society (SPS) Tunisia Chapter on January 2009, and he is actually the chair of this Chapter. During this period, the chapter organized five IEEE Distinguished Lectures and other technical and professional activities. He is the current advisor of the IEEE SPS Student Chapter in ENIS since 2010.

HESHAM ALHUMYANI received the Ph.D. degree from the University of Connecticut, Storrs, USA. He is currently working with the Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, where he was appointed as the Faculty Dean, in 2019. He has published many research articles in distinctive journals and conferences.

His research interests include wireless sensor networks, encryption, underwater sensing, the Internet of Things (IoT), and cloud computing.