

Received December 24, 2021, accepted January 12, 2022, date of publication January 25, 2022, date of current version February 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3145950

Audio Watermarking for Security and Non-Security Applications

MAHA CHARFEDDINE¹, (Member, IEEE), **EYA MEZGHANI¹**,
SALMA MASMOUDI¹, (Member, IEEE), **CHOKRI BEN AMAR²**, (Senior Member, IEEE),
AND HESHAM ALHUMYANI²

¹REGIM: REsearch Groups on Intelligent Machines, National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia

²Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

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


This work was supported by Taif University Researchers Supporting Project, Taif University, Taif, Saudi Arabia, under Grant TURSP-2020/216.

ABSTRACT The digitization of audiovisual data is significantly increasing. Thus, to guarantee the protection of the intellectual properties of this digital content, watermarking has appeared as a solution. 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 applications, where we have involved neural network architecture in the insertion and detection processes, and integrated some masking phenomena of the human psychoacoustic model with linear predictive coding spectral envelope estimation of the audio file. Experiments proved the efficiency of exploiting perceptual masking with spectral envelope consideration in terms of imperceptibility and robustness results. In addition, we suggest a second audio watermarking technique for non-security content characterization applications based on a deep learning classification architecture. In this scheme, the extracted watermark advises about the audio class: music or speech, speaker gender, and emotion. The reported results indicated that the suggested scheme achieved a higher performance at the classification level, as well as at the 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 for humanity. The advent of digitalization has led to a number of easy-to-use and reasonably cost-free channels for transferring ideas and exchanging information. Nonetheless, the instantaneous effect of digitization has been the proliferation of illegal copying that involves 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 opportunities for not only the protection of copyrighted data, but also serves as a general framework to embed information within generic data sorts

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung .

for various usages. In this paper, we explain several digital watermarking usages and classify them into security and non-security applications. Next, we introduce two digital watermarking techniques that 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 and explores its security and non-security applications with some previous works we have already developed in such fields. Section 3 presents the two proposed audio watermarking schemes for both security and non-security usage for standard and sensitive audio contents. For both distinctive schemes, comparison studies are established based on well-known state of the art approaches as well as recent works in each type of applications.

Finally, the conclusion is presented in the last section, along with perspectives for future research.

II. WATERMARKING DEFINITION AND APPLICATIONS

The watermarking system principally involves two parts: embedding and extraction processes. They generally use a cryptographic key, which can be a public or secret key. A watermark is the signature hidden in the original digital content. Watermarked documents are the output data resulted by superimposing the original document and signature. Watermark embedding is shown in Fig. 1, and the extraction process is shown in Fig. 2.

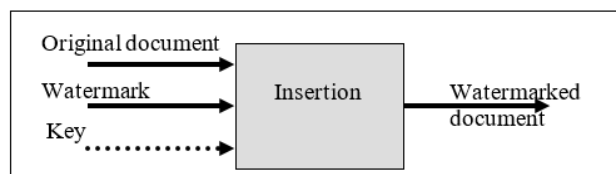


FIGURE 1. General digital watermark embedding process.

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

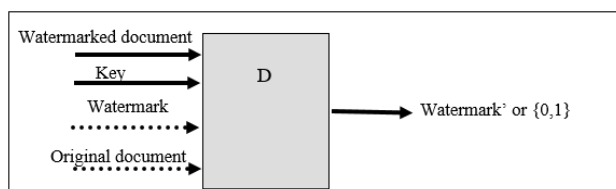


FIGURE 2. General digital watermark extraction process.

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

- Imperceptibility: This is the most important criterion in digital watermarking [8], [9]. However, we can retrieve in the literature some watermarking techniques that hide perceptible watermark [10].
- Robustness: This means that the hidden watermark in the data can endure different attacks and modifications. In most circumstances, we prefer watermarks to be robust [2], [11], [12]. In other cases, we wish for any processing in the watermarked document to jump the signature [13]–[15]. Finally, we wish an intermediate situation, where the mark persists in spite of some processing, and not for others, which we refer to a semi-fragile watermarking [16], [17].

- Security: Only legal users can extract the watermark; therefore, a proprietor can reach the goal of copyright protection.
- Capacity: This defines the maximum amount of data that can be hidden in a digital document. This capacity is habitually significant, as many systems need a large payload to be hidden.

Digital watermarking has multiple applications. In this study, we chose the following classification of watermarking applications:

A. SECURITY WATERMARKING APPLICATIONS

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

1) DATA HIDING

A well-known application where data are embedded and transmitted secretly in such a way that illicit person cannot discover it [18], [19].

2) SURREPTITIOUS COMMUNICATIONS

Principally, steganography is applied in the military applications, where people search to send secret messages to each other without being perceived [20], [21].

3) OWNERSHIP PROOF

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

4) AUTHENTICATION

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

5) PROPRIETOR IDENTIFICATION

It is somewhere written on the wrapper of an object, some information such as the identification brand of a paper maker. These kinds of watermarks can be effortlessly removed by cropping or tearing paper. To overcome this problem, watermark bits that identify the owner are hidden to form an integral part of the digital content [26], [27].

6) COPYRIGHT PROTECTION

The proprietor can hide the signature in the data to protect conspicuous content. There has always been a problem in supplying the owner identity of an object. In addition, if there is a disagreement concerning the data proprietorship, 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. The insertion and extraction processes depend on the back-propagation neural network architecture (BPNN). Furthermore, the choice of frequencies and the block covering the watermark were contingent on an earliest study of the effect of MP3 coding at different rates on the sound signal. Experiments showed that the proposed scheme presented good robustness and audio quality results. We also consider in the same paper [30] the adaptation of the proposed scheme in the video watermarking approach, which is different from our previous technique [31], which focuses only on video frames without considering the audio channel. In fact, in [30], we adjusted the MP3 study to a video watermarking scheme based on an earliest study of MPEG video coding. Once more, we achieve copyright protection purposes and ameliorate the robustness criteria of the video watermarking technique.

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

In section III, we will describe a novel audio watermarking scheme for copyright protection applications based on preliminary attacks, frequency masking studies, and spectral envelope estimation of basic and sensitive audio signals.

7) DIGITAL RIGHTS MANAGEMENT

This covers the mechanisms used by content publishers and rights holders to inflict access-licensing terms. They concern principally DRM for relational data and propose a set of watermarking techniques [33].

8) TRACEABILITY

Digital watermarking is used to trace the sender of a digital document copy [34]. The idea was 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].

In our previous works [36], we proposed a new watermarking based technique for traceability of multimedia documents. In this paper, we concluded that the tracing operation is frequently constrained by the absence of evidence regarding the number of colluders and also the collusion channels. The Tardos decoding operation is invariant regardless of the type of collusion, which can reflect its accusation performance. Thus, we proposed a new MAP-based estimation strategy to enhance the Tardos decoding step and

ensure a respectable estimation result. The proposed idea takes advantage of operating in a hierarchical context to deliver a more succinct and exact accusation decision in a short time. As a second tracing system [37], we proposed a confident fingerprinting approach based on a two-stage tracing strategy combining the Boneh Shaw and the replication schemes with the use of Tardos codes. This scheme was applied to a multilevel hierarchical fingerprint hidden using a DCT-based audio watermarking algorithm [38]. By taking advantage of grouping users and applying a weight-based tracing mechanism, the suggested fingerprinting technique reduces efficiently the computational costs of the tracing time and delivers a suitable solution that diminishes considerably the users recovery space and provides respectable robustness.

9) INTEGRITY VERIFICATION

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

We have also studied recently in [11] an integrity control application using an image watermarking scheme. This scheme extracts features from an original digital image to generate a watermark. To resist to rotation and cropping attacks, the technique uses Speeded-Up Robust Features [42] to localize invariant keypoints. Experiments prove that our scheme provides a high level of invisibility and robustness to standard JPEG compression and unique/double Stirmark attacks and achieves a successful level of integrity.

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 duplication recording is forbidden, the manufactured recorder needs to embrace the mark detection circuitry [43]–[45].

11) LOCATING DIGITAL CONTENT ONLINE

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

12) FORENSICS

This technique enhances the possibility of the proprietor detecting and responding to the abuse of its possessions.

It is exploited not only to gather proof for the criminal, but also to enforce the contractual usage agreement between the proprietor and the individuals whose digital content is shared with [47].

13) MEDICAL USAGES

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

B. NON-SECURITY WATERMARKING APPLICATIONS

In non-security watermarking applications, robustness to intentional attacks is not necessary and the watermark should generally contain a large amount of capacity information and must be extracted with a blind detection scheme. Among these applications, we found the following ones:

1) BROADCAST VERIFICATION

The aim of broadcast verification is to compile statistics on the use of digital content. Advertisers in radio broadcasts commonly want to guarantee that their announcements are correctly distributed according to the number of times specified in the contract. Therefore, a watermark was hidden in each advertisement. It permits, for example, to recognize in which radio the audio signal is broadcast, how many times, and at which moments [49].

2) MUSICAL EXTRACTS SEPARATION

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

3) INCREASING TELEVISION PROGRAMS INTELLIGIBILITY

This part focuses in replacing the teletext display in real time by inserting cloned versions into television programs. This will allow deaf and hard-hearing people to develop their comprehension skills based on the face and hands movement reproduced by the cued speech [51].

4) SOUND ANNOTATION

Sound annotation can be used to transfer labels to help in signal indexing. The embedded information can include metadata describing the signal content or information regarding a target application [52]. In this context, we introduced in [53] a watermarking scheme that performs multimodal video characterization and summarization. Audiovisual features were inserted as watermarks. Using the descriptors enclosed in the mark, key moments within a video, generally characterized

by high loudness or high motion, can be recovered easily by extracting the corresponding signature. Similarly, narrative video sequences, commonly known as low or medium motion loudness and activity, can be designated using a watermark. Besides, we can browse within the digital video, and we can extract scenes with particular properties, such as natural or artificial scenes and night or day views.

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

5) MOBILE USAGES

Digital watermarks offer an excellent opportunity for marketers to look for new behaviors to engage consumers with rich media experience on their phones. Watermarks can be easily hidden in all forms of media documents, including packaging, newspapers, posters, brochures, etc. [54]. This watermark is accorded to an URL in a backend database, which is consequently reverted to the consumer's smartphone.

6) MEASUREMENT OF AUDIENCE

Actually, audience measurement services should be reported more precisely and consistently from several channels. Watermarking hides a single identifier into a digital content for distribution and dissemination, which makes corresponding broadcasters quickly identifiable. The watermark provides 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 daily treatment and perfect reporting [55], [56].

C. WATERMARKING SENSITIVE DATA

As more communication and collaboration occur in the digital space, the requirement for maintaining data and document integrity is rising. Thus, businesses are attempting to increase their cloud security budgets. As an additional layer of security, they often choose to watermark digital documents when shared internally or externally. Watermarking helps in preventing recipients from data exfiltration activities and guaranteeing that sensitive informations (such as contracts budgets, confidential communication or manuscripts, health records) remain private and compliant during its lifecycle, so collaboration will be achieved with confidence.

For example, in the teleradiology context, the privacy and security of sensitive information have become serious issues [57], [58]. Teleradiology has been understood extensively as an eHealth service ended through remote diffusion of radiology information and images over electronic networks and the interpretation of the transferred images for diagnosis purposes. These radiology data, essentially Electronic Personal Health Information (EPHI), are exposed to potential alterations with severe complications, since they are very sensitive. Such information, needs to be protected and ensures high integrity and confidentiality.

Another example concerns the identity cards, which are also very sensitive and must be highly concerned. In fact, if the National ID card undergoes attacks, such as forged identity and counterfeit cards or content falsification, it will affect citizens and locate the issuing government in excruciating situations. A sensitive national identity card should include visible and invisible digital watermarks with secret text information [59].

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

The preservation of the sensitive nature of this type of data requires special digital watermarking algorithms, which represent challenges 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 that promise copyright protection of digital audio signals. Subsequently, we introduce our contributions to this type of audio watermarking application.

1) WATERMARKING TECHNIQUES RELATED TO COPYRIGHT PROTECTION APPLICATION

Copyright protection applications for digital content have become an essential issue. Digital watermarking techniques have received considerable attention to address this problem. This part presents a review of some papers focusing on copyright protection context.

In [63], authors 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). The experimental results specified that the hidden robust signature resist to attacks that are usually faced to. In addition, the system was resistant to cropping and replacement attacks. The perceptual evaluation confirmed that the watermark produced little degradation with an average Signal to Noise Ratio SNR [2] from 19.17 to 20.07 dB. Thus, inaudibility propriety should be ameliorated.

A new audio watermarking technique with good robustness was discussed in [64] by discovering the multi-resolution characteristics of the Discrete Wavelet Transform (DWT) and the energy compaction capability of the Discrete Cosine Transform (DCT). The watermark was embedded by slightly altering some frequencies of the audio signal. The audio fragments are segmented using DWT to obtain numerous groups of wavelet coefficients in several frequency bands. The fourth-level detail coefficients are then selected to be inserted in DCT domain to obtain two sets of transform domain coefficients. The average amplitudes of the two sets were modified to hide a binary image. The watermark detection is blind. The experimental results confirm that the suggested algorithm gives acceptable inaudibility performance with an SNR about 23.49 dB and a large capacity of insertion. From the obtained results against noise corruption attack, we notice that the quality of the detected watermark is altered. In fact, if the additional noise is important, the detection accuracy is reduced.

In paper [65], authors presented an audio watermarking technique in the DWT domain based on mean-quantization and using planar and binary images as signatures encrypted with chaos sequences. In this scheme, the audio file is segmented using a suitable wavelet basis. Low-frequency coefficients are used to hide watermarks with a mean quantization algorithm. The watermark can be detected without the needs of the original audio file. Comparison with known prior quantization watermark embedding schemes shows better robustness and good inaudibility results against different types of attacks.

In [66], a blind and adaptive audio watermarking technique was introduced based on a chaotic encryption scheme operating in the DCT and DWT hybrid domains. 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 increase efficiently the inaudibility and robustness results. The signature is encrypted using a chaotic sequence to enhance the watermark security. Experimental tests showed an acceptable inaudibility with an average SNR of 24.58 dB. This scheme ensured more security and good robustness faced to ordinary signal-processing attacks and not to malevolent operations.

A new blind watermarking technique is proposed in [67] exploring the auditory masking properties and the rational dither modulation (RDM) in the DWT domain. The insertion of binary information is assured by modulating coefficient vectors in the 5th-level approximation sub-band. The robustness and capacity of the suggested scheme can be controlled by changing the vector dimensions, whereas the inaudibility is guaranteed by constraining quantization noise under the auditory masking threshold. Besides, the periodic characteristics used in the RDM formulation can be exploited to re-ensure synchronization for truthful watermark extraction. Experiments exhibited that this technique supplied a near-zero objective difference grade and an average SNR close

to 20 dB. Compared to other developed techniques, the proposed scheme attained cognate robustness results.

In [68], a technique that inserts the watermark into the maximal DCT coefficients of a moving average sequence was proposed. In fact, signal processing operations generate noise that usually modifies the high frequencies of an audio file. Thus, hiding watermarks by regulating low-frequency coefficients can enhance the robustness of the watermarking algorithm. The moving-average sequence is a low-frequency feature of an audio file. Subjective and objective tests proved that the suggested watermarking technique preserves good audio quality and more resistance to the most known digital signal processing manipulations. However, comparison with other exiting schemes, measuring the performance of this algorithm, is limited to only low-pass filter processing.

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

2) INTRODUCTION OF THE WATERMARKING TECHNIQUE FOR COPYRIGHT PROTECTION APPLICATION

In this section, we present an enhanced approach to our previous proposed audio watermarking technique [2] based on DCT transform and a 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. This scheme presents the ability to be extended for content integrity and tamper detection applications. In this approach, we insert the watermark into the middle-frequency bands after performing the DCT transform. To improve the robustness and the security results, while maintaining good inaudibility performances, we exploited the BPNN architecture in the embedding and extraction processes [30]. The basic idea is to establish a relationship between the frequency samples around a central sample by using the BPNN model. In fact, for a selected transformed sample $I(x)$, the NN is trained using the eight neighbors as the input vector and the value of the sample as the output. The BPNN architecture covers three layers: an input layer with eight neurons, a hidden layer with nine neurons, and an output layer with a single neuron. After performing frame division of the original audio signal, the DCT transform was applied to the resulting frames. Next, each transformed frame was divided into nine samples to form a block, as shown in Fig. 3. The center sample of the block is the output and the neighboring samples are the input. Finally, we proceed to the NN training until a definite goal or specified maximum number of iterations is reached. When BPNN training is completed, a set of synaptic weights (w_i), characterizing the behavior of the trained network, can be obtained and used in the BPNN simulation of the embedding and extraction processes.

The originality of this new scheme resides in the exploitation of the perceptual masking frequency of the Human Psychoacoustic Model HPM [69] associated with the Linear Predictive Coding LPC [70] spectral envelope estimation of the digital audio file. In fact, after studying the HPM,

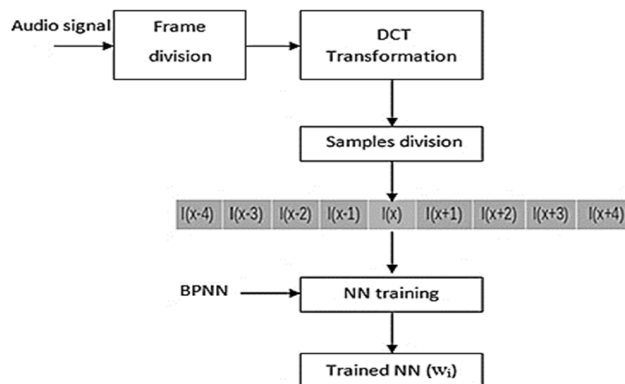


FIGURE 3. BPNN training process.

we obtained the masking threshold curve L_{tg} [69] and compared it with the LPC envelope to hide the watermark under this curve properly and imperceptibly. Another specificity of this scheme is the completely blind detection process, in contrary with previous schemes [2], [30], because neither the original audio signal nor the secure key is saved and transmitted to the receiver. In fact, the frame positions and corresponding indexes of insertion are recalculated in the detection process, which guarantees blindness. Experimental results showed that the exploitation of perceptual masking with 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 Stirmark attacks before watermarking. It permits to choose the adequate attacks that are suitable to the copyright protection context. The robustness of our scheme was evaluated again for the chosen attacks after the watermarking process. So, different MATLAB simulations were achieved. Table 1 displays a selection of the studied standard music signals and sensitive Quranic audio files. All signals have been sampled using a sampling frequency of 44.1 kHz, a number of bits per sample equal to 16bps and an approximate duration 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 used for measuring the robustness of our proposed watermarking technique, we are based

on one of the International Federation of Phonographic Industry IFPI requirements [30], which inflicts that the watermarking algorithm must avoid unauthorized removal of the hidden watermark unless the audio signal quality becomes very humble. Therefore, in the preliminary study, we applied all Stirmark attacks to the original audio signals to discern audio quality after attacks and to discard the attacks that corrupt the audio quality extremely. In fact, if an audio signal is very damaged, robustness will not be guaranteed. We computed the Signal to noise ratio SNR [2] values in decibels, between the original audio signals and the corresponding attacked ones.

Besides, to verify the quality of audio files, we achieved the Subjective Difference Grade SDG tests based on Recommendation UIT-R BS.1116 [8] and obtained the corresponding values and their descriptions.

Preliminary studies are presented in Tables 2, 3, 4 and 5.

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

When applying these attacks, we do not perceive degradation of the audio quality. All subjective results are often “imperceptible” with an SDG = 0 and sometimes “Perceptible but not annoying” with SDG = -1. The SNR values are positives, except for the attacks “invert” (number 5) and “fft_invert” (number 6) in Table 2. In fact, 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 imaginary values in the frequency domain of the sample values. Therefore, it is obvious to get negative SNR values while having no degradation in audio quality (SDG = 0, description=“imperceptible”). The attacks from 26 to 29 are audio manipulations that do not exist in Stirmark attacks. Conversion 16_8_16 (number 23) changes the number of bits per sample from 16 to 8, and vice versa. Cut_replace_samples_1, Cut_replace_samples_10, and Cut_replace_samples_20 (numbers 24, 25, and 26) are combinations of two attacks “cutsamples” (number 51) and “copysamples” (number 52) described in Table 5. For example, Cut_replace_samples_20 removed 20 samples every 1000 samples and replaced them with a new set of 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.

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

When applying these attacks, we observed distinct irregularities in the results.

Irregularity Type 1: At this point, we observe that the results vary from one signal to another. In fact, 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 that the results of the objective SNR test and the subjective SDG test are

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	
1	Stat1	21.39	0	Imperceptible	20.73	0	Imperceptible
1	Stat2	35.32	0	Imperceptible	29.68	0	Imperceptible
1	Re sample 44.1 32 44.1	62.81	0	Imperceptible	44.57	0	Imperceptible
1	Re sample 44.1 22.5 44.1	43.29	0	Imperceptible	27.95	0	Imperceptible
1	AddBrumm 100	37.05	0	Imperceptible	29.08	0	Imperceptible
1	AddBrumm 1100	16.18	0	Imperceptible	8.21	0	Imperceptible
2	AddBrumm 2100	10.56	0	Imperceptible	2.59	0	Imperceptible
2	Addnoise 100	39.40	0	Imperceptible	31.44	0	Imperceptible
2	Addnoise 300	29.82	0	Imperceptible	21.85	-1	Perceptible but not annoying
2	Conversion 16 8 16	30.63	0	Imperceptible	22.58	-1	Perceptible but not annoying
2	Cut_replace samples_1	91.54	0	Imperceptible	43.81	0	Imperceptible
2	Cut_replace samples_10	81.64	0	Imperceptible	43.82	0	Imperceptible
2	Cut_replace samples_20	78.56	0	Imperceptible	43.82	0	Imperceptible

not equivalent. For example, for the audio file “svega.wav” and the attack “addnoise_500”, we obtain 14.48 as SNR and “very annoying” as decision of the subjective test.

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	-	-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

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 low SNR value equals to -1.16.

For the Stirmark attacks part-2 presented in Table 4, we cannot expect 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. These attacks will be considered in our watermarking robustness tests.

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

Undoubtedly, we perceive a significant degradation of the audio quality face to these attacks. In fact, all subjective results are usually “Very Annoying” with an SDG = -4 for all original audio files. The SNR values present lower values except for the attack “addsinus” (number 43). Besides, attacks 49 to 52 are the worst attacks that remarkably affect audio quality. As a result, 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 for these attacks as the obtained attacked audio files

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 Flipsample	0.45	-4	Very Annoying	0.64	-4	Very Annoying
46 Fft_stat1	1.23	-4	Very Annoying	1.63	-4	Very Annoying
47 Addffnoise	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

are very different from the original ones (they do not have the same dimensions). As our proposed audio watermarking technique is typically used for copyright protection applications, we conclude that it is not interesting to study the attacks in Table 5 for the robustness tests.

In fact, applying these attacks to a 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 presented in table 5 which are “addsinus” (number 43), “echo” (number 44) and “flipsample” (number 45).

This choice is argued by the fact that it is possible for a pirate to apply them to remove the watermark without gathering that it will damage the auditory quality of the attacked watermarked signal. Furthermore, we combine two attacks and perceive their effects on the watermark. Because the attacks “cutsamples” (number 51) and “copysamples” (number 52) significantly destroy the audio quality if they are applied individually, we decided to combine them. We first removed one (or ten) (or twenty) sample (s) 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 were “Cut_replace_samples_1” “Cut_replace_samples_10” and “Cut_replace_samples_20”. We categorized the obtained attacks in Table 3 (numbers 24 to 26) as they always present a very high SNR and subjective results “imperceptible” with an SDG = 0.

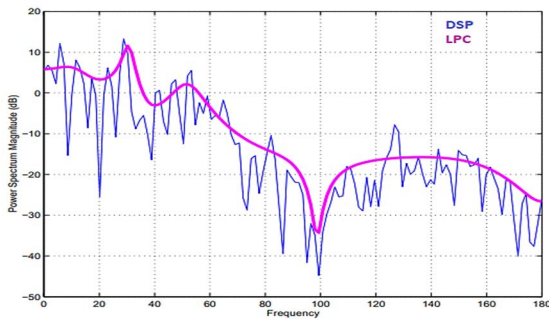


FIGURE 4. The smoothing aspect of LPC curve with minimum variations vs PSD curve.

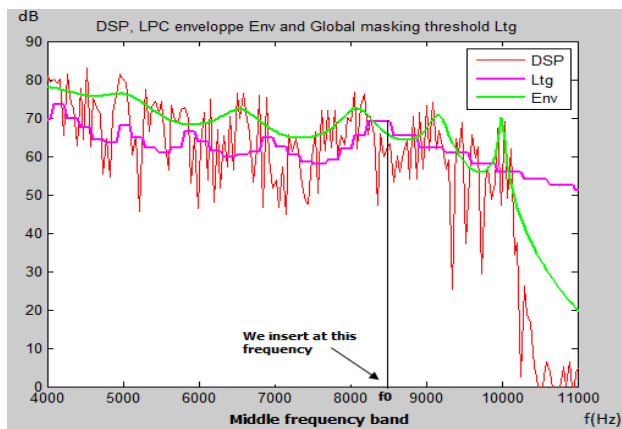


FIGURE 5. The estimated envelope Env compared with Ltg.

We explain in the following parts our proposed audio watermarking technique for copyright protection applications.

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

The MPEG audio standard [69] encodes an audio file by eliminating the acoustically irrelevant portions of the audio data. In reality, it benefits from the inability of the human auditory system to perceive quantization noise beneath auditory masking conditions. 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 the seventh HPM steps. HPM imposes that to have an imperceptible quantization noise, we should stay below the masking threshold curve. We attempted to create an analogy between compression and watermarking. Because the quantization noise resulting from the MPEG audio compression is inaudible when it is under Ltg, we anticipate then that the noise caused by watermark insertion will also be inaudible if it is under 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 HPM1 computes the PSD of 512 audio frames. These frames were overlapped by 128 samples as the 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” has a sample size of 384. These frames are used later to embed the watermark bits 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 Fig. 4. The LPC is universally used for sensitive envelope estimation and offers a smooth representation of important and delicate sound properties. The idea of LPC estimation is to represent each current audio sample $x(n)$ by a linear combination of its p prior values $x(n-p)$ through $x(n-1)$. p is the order of LPC [70]. Fig. 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, f_0 can be the adequate frequency at which we delicately insert 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, if $Ltg > Env$, then $diff_positive = Ltg - Env$
 We calculate next, the maximum difference from the computed positive differences “diff_positive”.

It is important to note that the localized middle-frequency band should be significantly narrow so that it will be the same calculated during the detection process. Consequently, the retrieved band, frames and insertion positions remain identical to those calculated in the embedding process. This, ensures the watermark resynchronization after attacks and signal processing.

- Finally, after accomplishing the three previous steps for all the overlapped 512 frames, we obtain N frequency values where we can embed the watermark. We necessary generated 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.

The embedding and detection processes of this scheme are defined in the following paragraphs.

- DCT-NN-HPM watermark embedding process

In the previous DCT-NN audio watermarking scheme, the audio signal was separated into non-overlapping frames of 512 samples and a DCT transform was achieved for each obtained frame. However, in the new DCT-NN-HPM, the provided non-overlapped frames “sframe” from the original audio division have a sample size of 384. Accordingly, the result is a DCT frame of 384 frequency samples size noted “sframe_DCT”. The obtained “sframe_DCT” was used next to cover the watermark bit.

In the preceding DCT-NN audio watermarking approach, we choose to hide the watermark bit in the middle-frequency band [4 kHz, 11 kHz]. For each frame and after localizing this band, we explored the sample value closest to the average value of the middle frequency located band and then deducted its position. The sample of the identified position covered the watermark bit. However, research of the insertion position 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 compute the positive differences in this band so that the LPC envelope was below Ltg. Subsequently, we calculate the maximum difference from the deliberated positive differences. The frequency sample corresponding to this maximum difference covers the watermark bit. The watermark insertion steps for the new DCT-NN-HPM are illustrated in Fig. 6.

- DCT-NN-HPM watermark extraction process

The DCT-NN-HPM detection process is displayed in Fig. 7. The searched frames and insertion positions constitute the proposed audio watermarking key. It is essential to note that this key is not transferred secretly to the receiver but is recalculated in the detection process, which implies complete blindness detection of the new technique in contrast to the previous technique. In fact, the searched frames and positions of insertion from an adequate narrow middle frequency band are the result of applying the first seven steps of HPM1 on the watermarked audio file. This stage is very significant in the detection process since it ensures the re-synchronization of de-synchronized frames and correspondent insertion positions in the case of de-synchronizing attacks. Thus, identical recalculated frames and embedding positions will be retrieved in the extraction process. 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 Fig. 7. We display in the following paragraphs the experimental results of 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 the compression robustness, we used the 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 used in the literature are the normalized cross-correlation NC [2], [30] and the Bit Error Rate BER [2], [72], [73]. They assessed the similarity between the extracted and the inserted watermarks. More NC is close to 1, more extracted watermark is similar to the embedded watermark. In the contrary, more BER is close to 0, more extracted watermark is similar to the hidden one.

In our tests, we assume that the watermark, which is a binary logo of size 32×32 , exists if the calculated correlation exceeds the threshold value of 0.7. In fact, if NC exceeds this threshold, the extracted watermark is perceptibly similar to the hidden watermark. Moreover, we consider that the watermark is correctly extracted if the computed Bit Error Rate value is less than 0.3. Effectively, if the BER is below this threshold, the detected watermark is perceptibly comparable to the embedded one.

The most well-known type of removal attack is the lossy compression. The common standard lossy compression for audio signals is MPEG 1 Audio Layer III MP3, which is regularly used in audio consumer storage. Different bit rates were used for the MP3 standard. 128 Kbps bit rate is usually used [74] at a compression ratio of 11:1, guaranteeing generally adequate sound quality. We tested the robustness of the proposed watermarking approach using three MP3 compression rates (128, 96 and 64Kbps). These chosen bit rates are the most frequently used rates in prior audio watermarking techniques [19], [41], [63], [64], [67], [75]–[77].

- Inaudibility results

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

Due to the exploitation of frequency perceptual masking related to the LPC estimation of the digital audio signal, the obtained SNR values, depicted in Fig. 8.a, are between 39 dB and 52 dB and are significantly higher than the value designed by the IFPI (20 dB).

Another well-known objective metric is used in Fig. 8-b and called Objective Difference Grade (ODG) [78], [79]. Closer the ODG value to 0, more degradation is not noticeable. The achieved ODG values confirm the watermark transparency of the proposed scheme.

- MP3 robustness results

Fig. 9 exhibits the MP3 robustness results. For all audio signals, we achieved very good MP3 robustness results (even, with a compression rate of 64Kbps, we obtain usually NC values greater than 0.87).

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

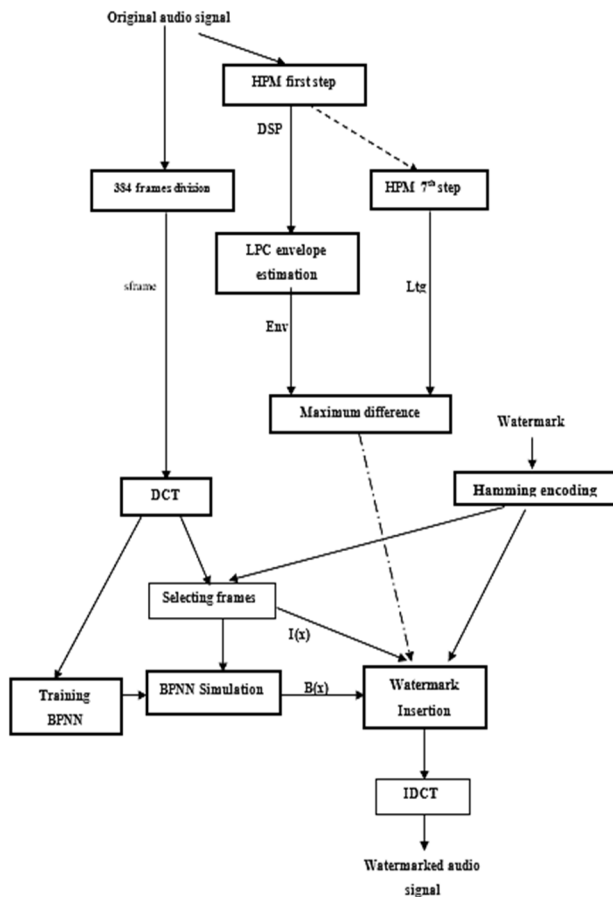


FIGURE 6. DCT-NN-HPM insertion process for security copyright protection application.

- Stirmark attacks part-1 robustness results

Fig. 10 presents the robustness results of Stirmark attacks part-1. We deduce that the DCT-NN-HPM scheme has good robustness, except for invert/fft_invert attacks.

- Stirmark attacks part-2 robustness tests

Fig. 11 displays the results of 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

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

We conclude finally that the experimental results reveal that the exploitation of frequency perceptual masking studied in HPM with spectral envelope estimation in the frequency domain is very interesting as we obtain very good inaudibility and robustness results.

6) INAUDIBILITY AND ROBUSTNESS COMPARISON WITH OTHERS

In this section, we exhibit comparison results using our DCT-NN based scheme [2] and other audio watermarking techniques. We give the BER, NC and SNR of different marks and audio files for all the compared schemes. Here, we consider close insertion capacity and we assume the average values of those metrics as in the literature survey.

- Inaudibility comparison with others

Exploring Table 6, we observe that the DCT-NN-HPM approach was the most efficient audio watermarking scheme in terms of inaudibility. Moreover, our previous scheme [2] and the technique in [80] assure also good imperceptibility results.

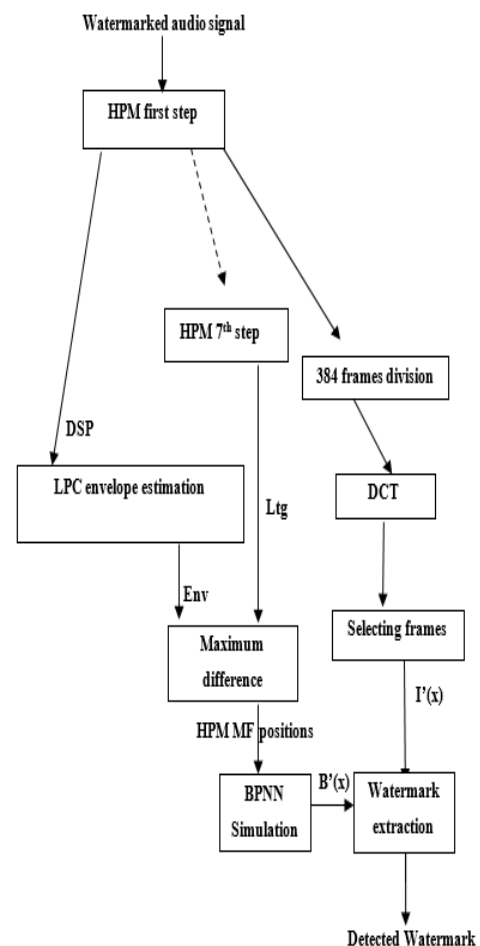


FIGURE 7. The DCT-NN-HPM extraction process for security copyright protection application.

- MP3 compression comparison with others

We compared the robustness to MP3 attack of the introduced audio watermarking technique with others. The results are presented in Tables 7 and 8, respectively. “X” means that the equivalent technique does not treat the indicated attack. When examining the compression results, we notice that our

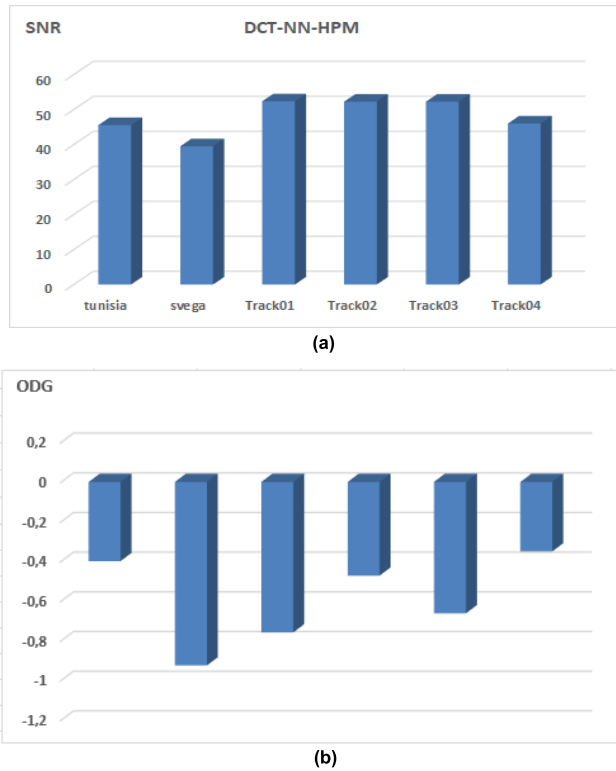


FIGURE 8. SNR/ODG values of the DCT-NN-HPM scheme.

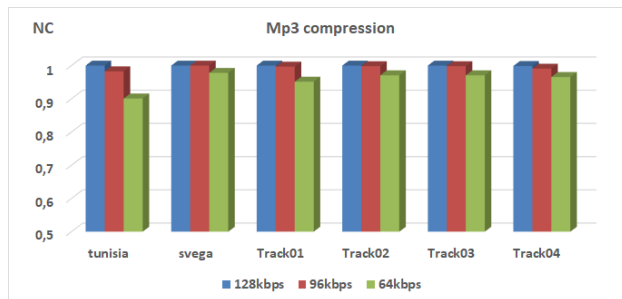


FIGURE 9. MP3 robustness results of the DCT-NN-HPM.

suggested technique DCT-NN-HPM presents the best results when using BER or NC metrics and considering the three compression bitrates. This observation proves the effectiveness of integrating the HPM masking study in the embedding algorithm. Besides, the schemes in [64], [66]–[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, which shows the comparative Stirmark attacks between our proposed scheme and others by using the normalized cross-correlation NC, we notice that our suggested scheme DCT-NN-HPM gives the best robustness results since all the NC values are 1 or very close to 1.

Moreover, if we observe Table 10, which shows the comparative Stirmark attacks between our suggested technique

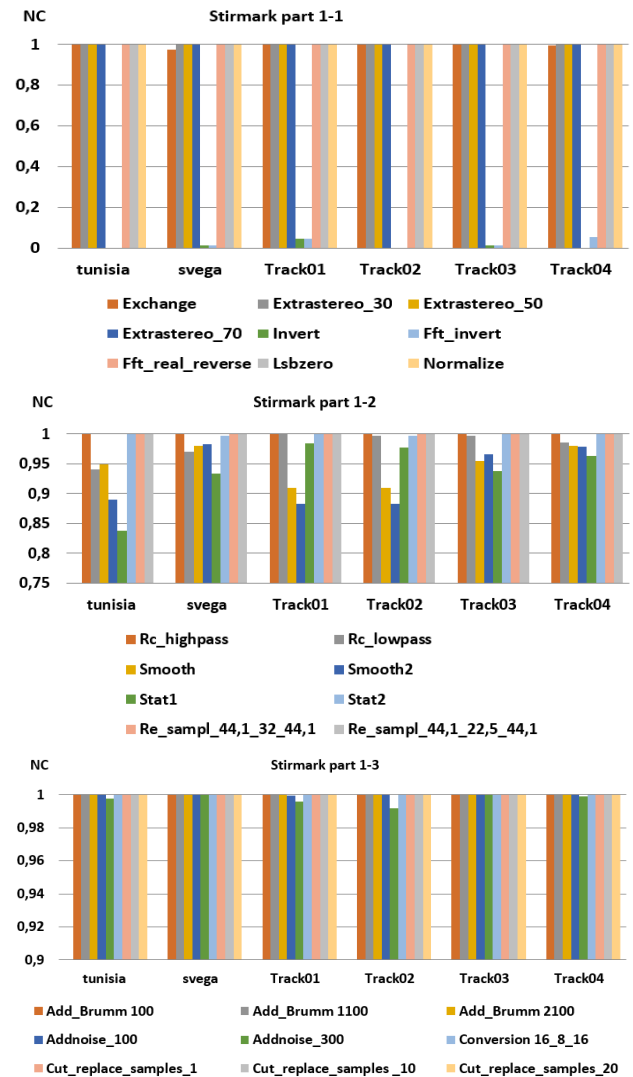


FIGURE 10. Stirmark attack part-1 results of the DCT-NN-HPM.

and existing works 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 close to 0, except for the invert attack.

In addition, the techniques in [2], [63]–[65] are resistant to Stirmark attacks.

In the following section, we describe a second audio watermarking scheme for non-security usage, focusing on audio content characterization and deep learning classification architecture.

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. Then, we introduce our suggested systems to this type of audio watermarking application.

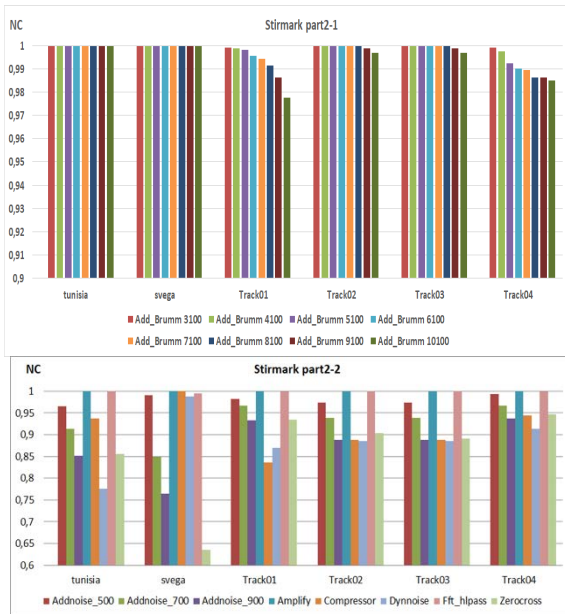


FIGURE 11. Stirmark attack part-2 results of the DCT-NN-HPM.

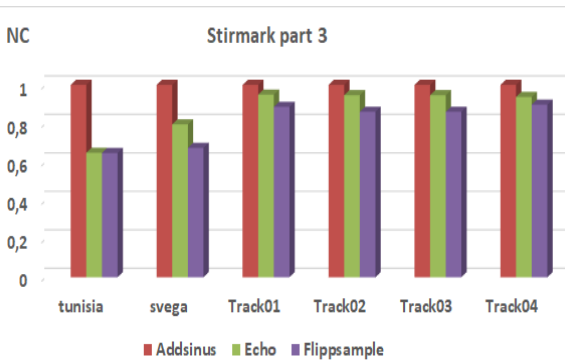


FIGURE 12. Stirmark attack part-3 results of the DCT-NN-HPM.

1) WATERMARKING TECHNIQUES RELATED TO CONTENT CHARACTERIZATION APPLICATION

Some prior techniques were debating content characterization by watermarking scheme.

In [92], the 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 that involve a series of sinusoids with harmonically allied frequencies. Thus, the authors considered the Short-Time Fourier transform (STFT) of the audio file to extract parameters that can be exploited to classify or watermark the signal. Hence, the authors suggest a new spread-spectrum watermarking algorithm using Instantaneous Mean Frequency (IMF) estimation of the original audio signal and simultaneous masking to obtain optimal points for watermark insertion. Results confirmed that the watermark was inaudible, statistically unnoticeable and robust to ordinary signal processing manipulations with BER 0-13%.

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
Lifting wavelet transform (LWT)-based framework [63]	20.07
Modifying the Average Amplitude in Transform Domain [64]	23.49
Compressive Sensing [80]	41.54
Wavelet-coefficients quantization [81]	21
Wavelet-coefficients Mean-quantization [65]	37.97
Asymmetric turbo-Hadamard code [82]	29.63
Singular-value decomposition [83]	25.24
Chaotic Encryption in Hybrid Domain [66]	24.58
Spread spectrum [84]	28.59
DC-level shifting [85]	21.24
Echo-data hiding [86]	21.47
Phase-coding [86]	12.2
Frequency-masking [87]	12.87
Empirical Mode Decomposition [88]	25.415
Fast Walsh Hadamard Transform [89]	33.83
Wavelet based technique [90]	32.45
DWT-based rational dither modulation [67]	20.21
Moving Average and DCT [68]	30.93
Air Channel Characteristics [91]	>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 [81]	X	X	0.84
Wavelet-coefficients Mean-quantization [65]	X	X	0.77
Asymmetric turbo-Hadamard code [82]	1	X	X
Chaotic Encryption in Hybrid Domain [66]	1	X	0.99
Fast Walsh Hadamard Transform [89]	X	X	0.99
Wavelet based technique [90]	0.97	X	X

To consider other type of attacks, authors in [92] proposed to ameliorate their approach by considering additional operations as redundancy coding or bits synchronization.

In [93], a TF-based audio coding algorithm with a new psychoacoustic model, music classification, audio classification, audio fingerprinting, and audio watermarking was introduced to demonstrate the benefits of using time-frequency methods to study and extract information from audio files. The authors used the IMF estimation of the audio signal and nonlinear TF signature as marks. They proposed chirp-based watermarking, in which linear phase signals are hidden as TF signatures. To compensate the BERs in the estimated watermarked audio signal, the Hough-Radon transform (HRT) is used as a chirp detector in the post-processing process. This technique could correct the error up to BER of 20% and its robustness was acceptable. It offers higher BER correction than the repetition and the Bose-Chaudhuri-Hocquenghem BCH coding [93].

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
Lifting wavelet transform (LWT)-based framework [63]	0.04	X	10.10
Modifying the Average Amplitude in Transform Domain [64]	0.01	X	0.08
Wavelet coefficients quantizing [81]	X	X	0.23
Wavelet-coefficients Mean-quantization [65]	X	X	0.29
Singular-value decomposition [83]	0	X	X
Chaotic Encryption in Hybrid Domain [66]	0.01	X	0.06
Fast Walsh Hadamard Transform [89]	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]	[82]
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

In [94], the authors used state-of-the-art frame selection to suggest a new approach to preserve most of the discriminative features of speakers and to safe speech signals by applying the speech watermarking method. Thus, linear predictive analysis was exploited for each frame to extract the gain, formants, and residual errors. Consequently, a frequency-weighted function was utilized to quantify the formants, and high-order correlation with error gain was exploited for weighting the residual errors. Experiments revealed an overall 12% efficiency value in terms of performance, memory and time of frame selection for speaker recognition and speech watermarking approaches.

Paper in [95] presented a technique for joining biometric speech authentication and watermarking to assimilate metadata into the authentication process, which lacks important

TABLE 10. Comparative stirmark attacks (BER) of the DCT-NN-HPM with others.

Attacks	DCT-NN-HPM	[2]	[63]	[64]	[65]	[66]	[84]	[91]
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

quality and performance damages. Different audio watermark schemes were introduced to hide metadata as supplementary information into the reference data for 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 of the selected technique in the context of the gender belonging of 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 and these depended on the applied watermarking technique. Thus, authors need to consider the inconveniently insertion behavior of some watermarking techniques as they are not capable to hide the whole metadata. Consequently, they need to examine the potential application fields and the essential metadata to define the watermarking properties as well as the requisite recognition precision.

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 and ranging from music to speech. Therefore, if speech analysis is based on phonemes lasting a few tens of milliseconds, the 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 tolerable for a particular classification problem, it is entirely unacceptable for other problems. Based on the hypothesis that humans are the best generalist classifiers of audio signals, the proposed approach suggests inspiring human perception mechanisms in order to develop automatic audio classification systems. We propose a model of hearing memory and a feature set for psychoacoustic inspiration. Motivated by the great development of deep learning at the expense of classic learning algorithms, we propose to combine the feature vector with Deep Neural Networks to develop an audio classification system. The 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. The experimental results are reported exhibiting the retrieved performance on public datasets. Finally, the robustness and transparency of the watermark were assessed.

We notice that for non-security watermarking applications, robustness to intentional attacks such as dy-synchronisation treatments is not necessary. Good robustness against licit signal processing as compression is required. In such applications, the watermark should commonly contain a great capacity information and must be extracted using a blind detection approach. One of the most popular applications of data transmission is sound document annotation. In fact, as we know, this application can be used to transfer labels to facilitate signal indexing. The inserted information contains metadata describing the signal content or information regarding a target application. For example, a hidden watermark can indicate the name of the artist, the place of registration, or any other data relating to the signal, as 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 audio content into audio classes: music, speech, male speaker, happy speaker, etc. So a watermark containing the information characterizing the audio content was constructed. In the extraction process, this watermark will inform about the audio class: music or speech, speaker gender, etc.

3) PROPOSED SCHEME FOR CONTENT CHARACTERIZATION USING AUDIO WATERMARKING

Fig. 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 more informative when conferring to the considered application. Audio files are usually divided into

overlapping windows. Next, descriptors were calculated for each frame. Statics are later made in longer-term windows. So, we can define two processing levels as displayed in Fig. 14: short-term and mid-term levels. Feature extraction, which 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 perform directly on the original signal, feature extraction can be viewed as a data amount reduction procedure. In order to get a higher accuracy, it is imperative to select the most appropriate feature set for the specific applications.

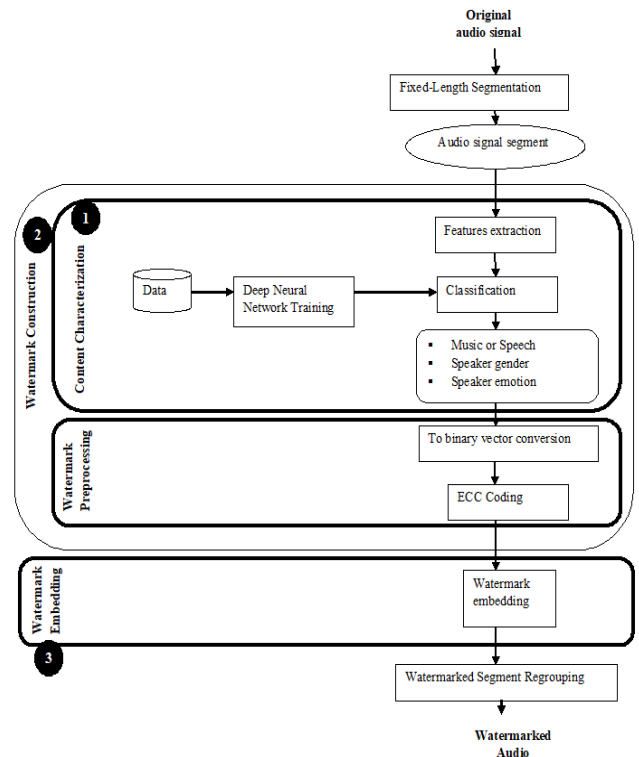


FIGURE 13. Proposed watermarking scheme for content characterization application.

-Short-term analysis

In short-term analysis, known as frame-based processing, the audio file is divided into overlapping frames, as exhibited in Fig. 14. The window duration at this level is approximately 10–50 ms, within which the signal is considered to be stationary. Consequently, descriptors can be extracted and computed during it [96]. After the framing step, windowing is typically applied to each frame to evade discontinuities at the block boundaries. In our approach, the Hamming window is selected at this step. After windowing, the deliberated features are calculated per frame, as presented in Fig. 14. As stated by the computational way, the extracted descriptors can be classified into time-domain and frequency-domain features.

Temporal Audio Features: These features were calculated directly from the audio samples. The most well-known time-domain features are Short-term energy [97], [98], energy

entropy and Zero-crossing rate [97]. These features will be utilized in the feature extraction stage of our technique because they guarantee a simple and good means for audio signal analysis.

Spectral audio features: In order to guarantee correct audio analysis, it is essential to combine time-domain and frequency-domain features, also called spectral features. These metrics were computed using Discrete Fourier Transform (DFT) coefficients of the designed audio frame. The most known spectral-domain features are spectral flux, spectral centroid, [99], spectral roll off, Mel-Frequency Cepstrum Coefficients (MFCCs) [97], [99], chroma vector [100] and Relative Spectral Analysis-Perceptual Linear Prediction (Rasta PLP) [101].

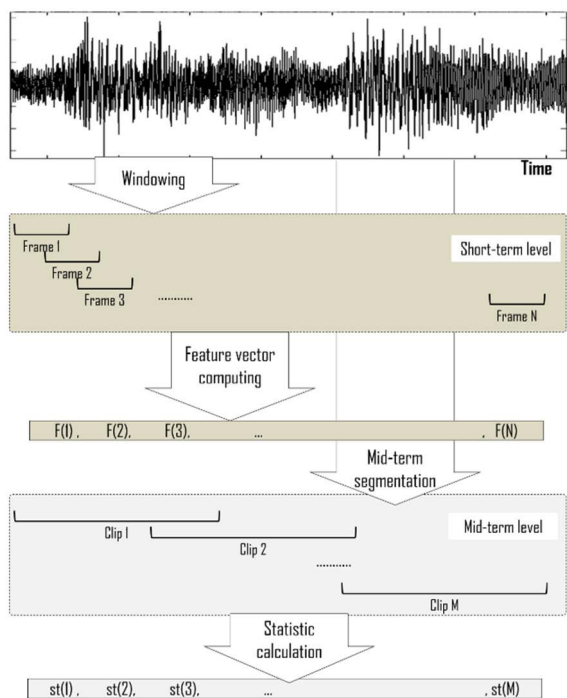


FIGURE 14. Audio signal decomposition.

-Mid-term analysis

After performing the short-term, known as 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 to be more appropriate. Later, 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 segments fixed between 1 and 10 seconds. Clips represent a set of successive frames and depict the behavior of short-term features. 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 Fig. 14.

In this paper, we consider four mid-term statistics: mean value, standard deviation, skewness and kurtosis [102].

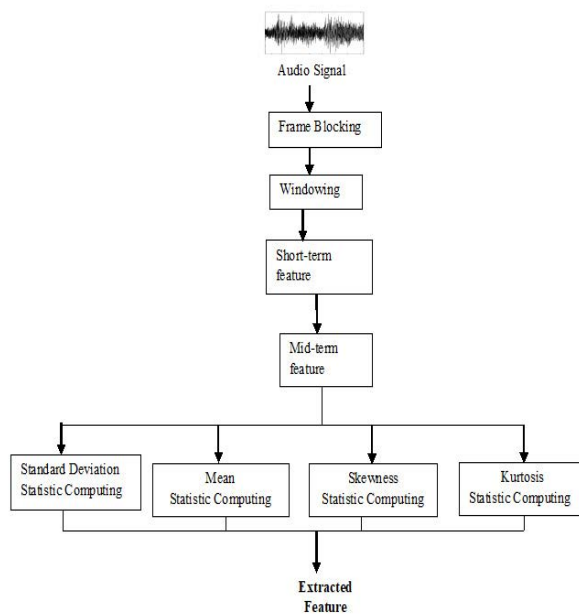


FIGURE 15. Fusion at feature level.

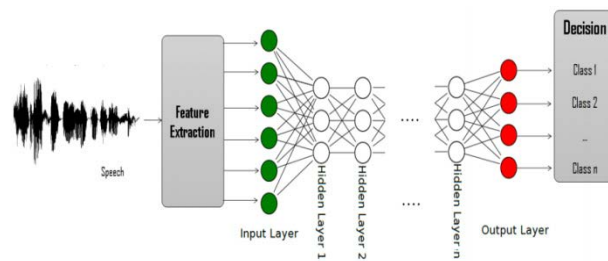


FIGURE 16. MLP deep neural network based audio classification scheme.

At first, each statistic metric was computed alone. After that, a fusion at the feature level is proposed and performed between the proposed statistics, as shown in Fig. 15.

- 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 Fig. 16. Block (1) of Fig. 17 targeting content characterization is performed using an MLP-based architecture for audio classification for three classification tasks: music and speech discrimination, speaker gender recognition, and speech emotion identification. Categorical cross-entropy is used as the loss function, and Softmax is used as the activation function for the last dense layer.

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

The proposed watermarking technique, DCT-MLP-LSB as illustrated in the Fig. 17 serves to characterize the host audio document. Indeed, at each segment, the detected watermark will inform about the audio class: music or speech, speaker

gender, etc. We start by detailing the watermark construction block and then move to the mark hiding process [5]. The original file was first split into a fixed-length segment. Each segment is analyzed and classified into audio classes: music, speech, male speakers, happy speakers, etc. Then, the retrieved information characterizing the audio content is inserted into the same signal. A binary vector is constructed using this information as follows: 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, the audio signal was divided into fixed-length blocks with 512 samples. Each block was transformed in the spectral domain by using DCT. The embedding region was selected in the middle-frequency band. The mean DCT value of this band was computed. The nearest frequency to this mean value was elected as the insertion position. The Least Significant Bit LSB at this position is then replaced by the watermark bit value. An inverse DCT was then applied. This process was repeated for each block along the audio stream. Therefore, each watermarked audio segment holds information about its content: music, speech, male speaker, speech emotion, etc. Detecting the watermark allows us to get these data and point to a moment according to a given criterion. The watermark detection process is the inverse embedding one. It begins by dividing the signal. Each segment is sliced into fixed length blocks with 512 samples. DCT Transform is applied. The LSB of the embedding position is selected. The watermark is constructed by Hamming decoding of the retrieved binary vector.

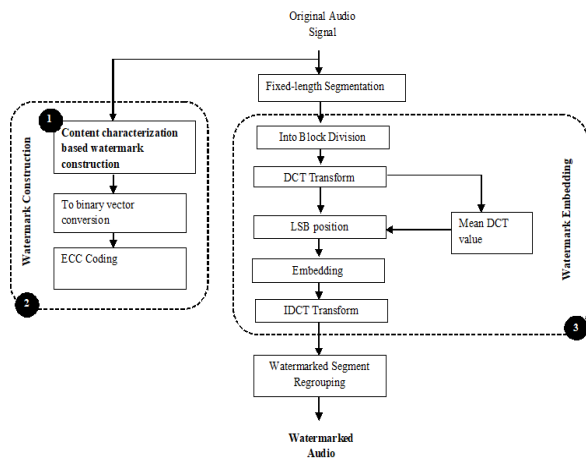


FIGURE 17. Proposed watermarking scheme of audio watermarking for deep learning based audio content characterization applications.

4) EXPERIMENTAL RESULTS OF THE NON-SECURITY WATERMARKING APPLICATION

- Classification assessment

In the next subsections, the 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

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
[107]	96.75%
[99]	93.5%
[108]	94.5%
[109]	95.9%
[110]	97.22%
[111]	97.28%
Our work on GTZAN	100%
Our work on S&S	100%

a collection of speech tracks and music excerpts assembled for classification purposes [103]. This involved 128 extracts lasting 30 seconds. These are mono 16-bit audio wav files sampled at 22050 Hz. This dataset comprises various musical styles and speech tracks that are recorded under

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
[112]	95%
[113]	98.65%
[114]	90.1%
Our Work-Eustace dataset	100%
Our work-Berlin dataset	98.9%

different conditions. The second corpus is Scheirer-Slaney (S&S) Music/Speech dataset [104]. It consists of a collection of 246 audio files saved in the WAVE format and during 15 seconds each one. These extracts were collected at random from the radio including music and speech. The experimental results for the two databases are reported in Table 11. The best

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%

achievement for the two datasets was attained by fusion of all statistics and using 50 neurons. The standard deviation outperforms other statistics when undertaken without fusion in all cases for the two datasets. The achieved performances of the prior approaches are depicted in Table 12. We notice that the suggested scheme attains higher performance.

- Experiments on speaker gender identification

The proposed system was experimented using two datasets from different languages: Eustace in English [105] and Berlin

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 %
	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%		

in German [106]. According to Table 13, gathering all statistics allows the enhancement of the classification accuracies. In fact, for the first dataset, the best achievement is obtained in the case of statistics fusion besides in case of computing one statistic standard deviation or mean value. Unlike the first database, the highest performance for the second dataset was achieved when all mid-term level statistics were fused

TABLE 17. Comparison of accuracy results for emotion speech recognition with previous work.

Ref	Best Acc Rate
[115]	49%
[116]	53%
[117]	72.05%
[118]	71.7%
[119]	76.3%
Our work-Berlin Dataset	61.9%
Our work-SAVEE Dataset	75%

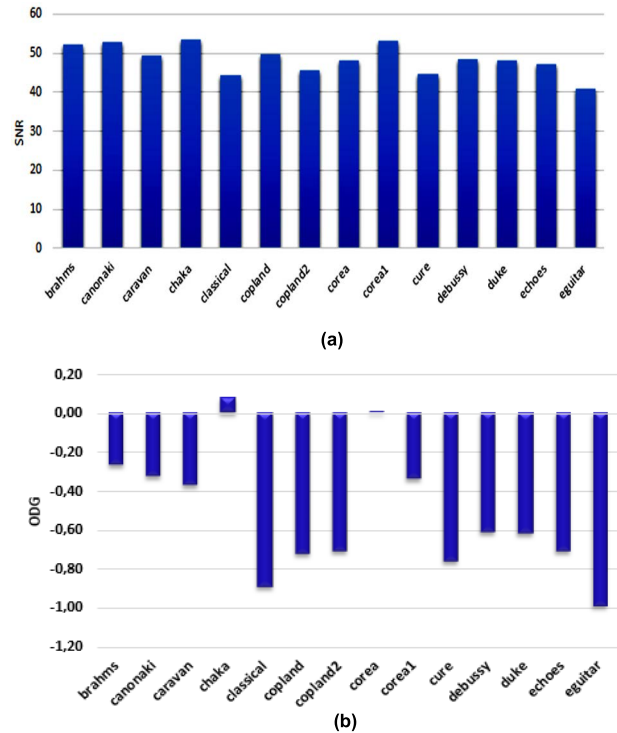


FIGURE 18. SNR/ODG values of the DCT-MLP-LSB scheme.

and 10 neurons were used for the three hidden layers. Unlike the first task, the mean value outperformed the other single statistics for this task. The performance achieved in some previous studies is reported in Table 14. It could be confirmed according to this table that the proposed scheme outperforms state-of-the-art approaches and affords promising results.

- Experiments on speaker emotion recognition

Two public datasets were used: the Berlin Database of Emotional Speech and Surrey Audio Visual Expressed Emotion (SAVEE) database. In order to evaluate system performance, the accuracy for each affective state is reported in Tables 15 and 16. The best rate was obtained in all cases of neuron numbers when using all mid-term level statistics, and the highest values were achieved in the case of 100 neurons for both databases. In the case of a single statistic, the highest performance is achieved in the case of the mean value for the first database, whereas the standard deviation outperforms other statistics for the second dataset. From Table 17,

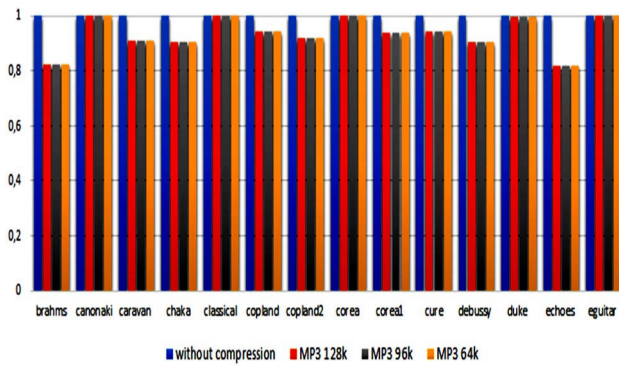


FIGURE 19. NC values after compression attacks in the DCT-MLP-LSB scheme.

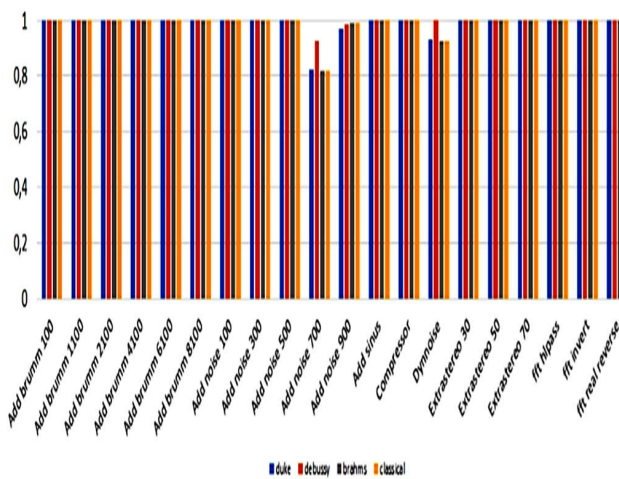


FIGURE 20. NC values after StirMark attacks in the DCT-MLP-LSB scheme.

we notice that the suggested technique achieves promising recognition rates compared to the state-of-the-art techniques.

- Watermarking evaluation

After the audio analysis process assessment, the watermarking algorithm was evaluated in terms of transparency and robustness.

- Watermarking transparency results

The Signal to Noise Ratio SNR, comparing the original and the watermarked files, was computed. According to the recommendation of the IFPI, transparency is confirmed as the SNR values exceed 20 dB. The results are presented in Fig. 18-a. According to the achieved SNR results, the watermark transparency was ensured by values more than 40 dB in all cases.

In addition, we observe in Fig. 18-b that all corresponding ODG values are close to 0. Thus, the DCT-MLP-LSB scheme satisfies the inaudibility requirements of an optimal audio watermarking technique.

- Watermarking robustness results

Since any signal is compressed before storage or transmission, the watermarking scheme should resist such

a transformation, and the watermark should be correctly detected even after compression. The MP3 encoder was experimented using a typical compression ratio since it is the most utilized audio encoder. As shown in Fig. 19, the NC values are higher than 0.8, confirming that the mark is almost detected. The robustness against StirMark attacks was then assessed, as shown in Fig. 20. NC values are equal to 1 in most cases, confirming the robustness of the proposed watermarking algorithm to the majority of attacks, except for the cases of Add noise 700 and Dynnoise attacks where the NC is slightly lower than 1. This problem can be circumvented by mark duplications.

IV. CONCLUSION

Digital audio watermarking can be used in different types of applications that target two different situations: the first for security applications and the second for non-security applications. Thus, in this paper, we carried a big attention in examining these situations. We then proposed two digital watermarking schemes that we implemented for basic and sensitive digital content. The first scheme was an audio watermarking technique for security copyright protection applications. This first work hides the signature in the narrow middle-frequency band of an audio frame. We have involved the NN architecture in the proposed insertion and detection processes to improve security and robustness, even with high watermark capacity. Furthermore, we studied and integrated some masking phenomena of the HPM. The objective was to determine the masking threshold curve and compare it with the estimated Power Spectrum Density envelope to appropriately insert the signature under this curve. Experimental results have proved that using frequency perceptual masking with spectral envelope estimation in the frequency domain offers good robustness compared with our previous NN-based audio watermarking technique [2] and with other existing watermarking techniques. In summary, we endorse that we have implemented an audio watermarking scheme that meets 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 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 suggest a second new audio watermarking approach for content characterization as non-security application. The originality consists of using watermark-holding information to characterize the audio content. Once detected, the user can browse the audio file and move to a selected moment according to the given criteria. For example, speech segments uttered by a male speaker with a happy emotional state can be picked out. For audio content analysis and classification, a deep-learning-based scheme was adopted and combined with a rich descriptor set. Moreover, for watermarking, a frequency-domain technique based on the DCT transform was employed. The reported results showed that

the proposed scheme achieves higher performance at the classification level as well as at watermarking.

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

REFERENCES

- [1] S. G. Rizzo, F. Bertini, and D. Montesi, "Fine-grain watermarking for intellectual property protection," *EURASIP J. Inf. Secur.*, vol. 2019, no. 1, pp. 1–20, Dec. 2019, doi: [10.1186/s13635-019-0094-2](https://doi.org/10.1186/s13635-019-0094-2).
- [2] M. Charfeddine, M. El'arbi, M. Koubaa, and A. C. Ben, "DCT based blind audio watermarking scheme," *Proc. Int. Conf. Signal Process. Multimedia Appl. (SIGMAP)*, Athens, Greece, 2010, pp. 139–144.
- [3] V. Bhat and G. I. D. A. Sen, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digit. Signal Process.*, vol. 20, no. 6, pp. 426–436, 2010.
- [4] M. Charfeddine, M. El'arbi, and C. Ben Amar, "A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain," in *Proc. 3rd Int. Symp. Commun., Control Signal Process.*, Malta, U.K., Mar. 2008, pp. 1138–1143.
- [5] M. Charfeddine, S. Masmoudi, M. Bellaaj, and C. Ben Amar, "Un schéma aveugle de tatouage audio numérique opérant sur les bits les moins significatifs dans le domaine fréquentiel utilisant un code correcteur d'erreurs," in *Proc. 6èmes Ateliers de Traitement et Analyse de l'Inf., Méthodes et Appl. (TAIMA)*, Hammamet-Tunisie, Tunisia, 2009, pp. 371–377.
- [6] Y. Terchi and S. Bougezuel, "A blind audio watermarking technique based on a parametric quantization index modulation," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25681–25708, Oct. 2018.
- [7] B. Dappuri, M. P. Rao, and M. B. Sikha, "Non-blind RGB watermarking approach using SVD in translation invariant wavelet space with enhanced grey-wolf optimizer," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31103–31124, Nov. 2020.
- [8] Z. Piotrowski and P. Gajewski, "Fidelity estimation of watermarked audio signals according to the ITU-R BS.1116-1 standard," *Acta Phys. Polonica A*, vol. 121, no. 1A, pp. A-82–A-85, Jan. 2012, doi: [10.12693/APhysPolA.121.A-82](https://doi.org/10.12693/APhysPolA.121.A-82).
- [9] S. Masmoudi, M. Charfeddine, and C. Ben Amar, "A robust audio watermarking technique based on the perceptual evaluation of audio quality algorithm in the multiresolution domain," in *Proc. 10th IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2010, pp. 326–331, doi: [10.1109/ISSPIT.2010.5711803](https://doi.org/10.1109/ISSPIT.2010.5711803).
- [10] T.-Y. Liu and W.-H. Tsai, "Generic lossless visible watermarking—A new approach," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1224–1235, May 2010.
- [11] N. Tarhouni, M. Charfeddine, and C. Ben Amar, "Novel and robust image watermarking for copyright protection and integrity control," *Circuits, Syst., Signal Process.*, vol. 39, no. 10, pp. 5059–5103, Oct. 2020.
- [12] R. Lancini, F. Mapelli, and S. Tubaro, "A robust video watermarking technique for compression and transcoding processing," in *Proc. IEEE Int. Conf. Multimedia Expo*, Lausanne, Switzerland, Aug. 2002, pp. 549–552.
- [13] X. Quan and H. Zhang, "Perceptual criterion based 'fragile audio watermarking using adaptive wavelet packets,'" in *Proc. Pattern Recognit. ICPR*, Cambridge, U.K., vol. 2, 2004, pp. 867–870.
- [14] M. Zamani and A. B. A. Manaf, "Genetic algorithm for fragile audio watermarking," *Telecommun. Syst.*, vol. 59, no. 3, pp. 291–304, Jul. 2015.
- [15] L. Rakhmawati, W. Wirawan, and S. Suwadi, "A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, p. 61, Dec. 2019.
- [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," *Int. J. Comput. Math.*, vol. 90, no. 12, pp. 2588–2602, Dec. 2013.
- [17] Z. Su, L. Chang, G. Zhang, J. Jiang, and F. Yue, "Window switching strategy based semi-fragile watermarking for MP3 tamper detection," *Multimedia Tools Appl.*, vol. 76, no. 7, pp. 9363–9386, Apr. 2017.
- [18] D. Taranovsky, "Data hiding and digital watermarking," in *Handbook of Visual Display Technology*, J. Chen, W. Cranton, and M. Fihn, Eds. Berlin, Germany: Springer, 2012.
- [19] A. S. Brar and M. Kaur, "A survey of reversible watermarking techniques for data hiding with ROI-tamper detection in medical images," in *Mobile Communication and Power Engineering (Communications in Computer and Information Science)*, vol. 296, V. V. Das and Y. Chaba, Eds. Berlin, Germany: Springer, 2013.
- [20] Y. Xiong and Z. X. Ming, "Covert communication audio watermarking algorithm based on LSB," in *Proc. Int. Conf. Commun. Technol.*, Guilin, China, Nov. 2006, pp. 1–4.
- [21] F. Li, B. Li, Y. Huang, Y. Feng, L. Peng, and N. Zhou, "Research on covert communication channel based on modulation of common compressed speech codec," *Neural Comput. Appl.*, vol. 809, pp. 1–14, Apr. 2020.
- [22] A. Deshpande and J. Gadge, "New watermarking technique for relational databases," in *Proc. 2nd Int. Conf. Emerg. Trends Eng. Technol.*, Nagpur, India, 2009, pp. 664–669.
- [23] T. Khanam, P. K. Dhar, S. Kowsar, and J.-M. Kim, "SVD-based image watermarking using the fast Walsh-Hadamard transform, key mapping, and coefficient ordering for ownership protection," *Symmetry*, vol. 12, no. 1, p. 52, Dec. 2019.
- [24] L. Priya C. V. and N. Raj N. R., "Digital watermarking scheme for image authentication," in *Proc. Int. Conf. Commun. Signal Process. (ICCSPP)*, Chennai, India, Apr. 2017, pp. 2026–2030.
- [25] A. Anand and A. K. Singh, "Watermarking techniques for medical data authentication: A survey," *Multimedia Tools Appl.*, vol. 80, pp. 1–33, Apr. 2020.
- [26] Y.-H. Chen and H.-C. Huang, "Coevolutionary genetic watermarking for owner identification," *Neural Comput. Appl.*, vol. 26, no. 2, pp. 291–298, Feb. 2015.
- [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," in *Proc. 7th Int. Workshop Biometrics Forensics (IWBF)*, Cancun, Mexico, May 2019, pp. 1–6.
- [28] S. Samuel and W. T. Penzhorn, "Digital watermarking for copyright protection," in *Proc. IEEE Africon. 7th Africon Conf. Afr.*, Gaborone, Botswana, Sep. 2004, pp. 953–957.
- [29] F. Ernawan and M. N. Kabir, "An improved watermarking technique for copyright protection based on tchebichef moments," *IEEE Access*, vol. 7, pp. 151985–152003, 2019.
- [30] M. Charfeddine, M. El'arbi, and C. Ben Amar, "A new DCT audio watermarking scheme based on preliminary MP3 study," *Multimedia Tools Appl.*, vol. 70, no. 3, pp. 1521–1557, Jun. 2014.
- [31] M. El'Arbi, M. Koubaa, M. Charfeddine, and C. Ben Amar, "A dynamic video watermarking algorithm in fast motion areas in the wavelet domain," *Multimedia Tools Appl.*, vol. 55, no. 3, pp. 579–600, Dec. 2011.
- [32] N. Tarhouni, M. Charfeddine, and C. Ben Amar, "A new robust and blind image watermarking scheme in frequency domain based on optimal blocks selection," in *Proc. CSRN*, 2018, pp. 78–86.
- [33] M. Zhaofeng, H. Weihua, and G. Hongmin, "A new blockchain-based trusted DRM scheme for built-in content protection," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–12, Dec. 2018.
- [34] F. Chaabane, M. Charfeddine, and C. Ben Amar, "A survey on digital tracing traitors schemes," in *Proc. 9th Int. Conf. Inf. Assurance Secur. (IAS)*, Dec. 2013, pp. 85–90, doi: [10.1109/ISIAS.2013.6947738](https://doi.org/10.1109/ISIAS.2013.6947738).
- [35] J. Franco-Contreras and G. Coatrieux, "Databases traceability by means of watermarking with optimized detection," in *Digital Forensics and Watermarking (Lecture Notes in Computer Science)*, vol. 10082, Y. Shi, H. Kim, F. Perez-Gonzalez, and F. Liu, Eds. Cham, Switzerland: Springer, 2017.
- [36] F. Chaabane, M. Charfeddine, W. Puech, and C. Ben Amar, "Towards a blind MAP-based traitor tracing scheme for hierarchical fingerprints," in *Neural Information Processing (Lecture Notes in Computer Science)*, vol. 9492, S. Arik, T. Huang, W. Lai, and Q. Liu, Eds. Cham, Switzerland: Springer, 2015.
- [37] F. Chaabane, M. Charfeddine, W. Puech, and C. Ben Amar, "A two-stage traitor tracing scheme for hierarchical fingerprints," *Multimedia Tools Appl.*, vol. 76, no. 12, pp. 14405–14435, Jun. 2017.
- [38] F. Chaabane, M. Charfeddine, W. Puech, and C. Ben Amar, "A QR-code based audio watermarking technique for tracing traitors," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 51–55, doi: [10.1109/EUSIPCO.2015.7362343](https://doi.org/10.1109/EUSIPCO.2015.7362343).
- [39] G. Zhang, L. Kou, L. Zhang, C. Liu, Q. Da, and J. Sun, "A new digital watermarking method for data integrity protection in the perception layer of IoT," *Secur. Commun. Netw.*, vol. 2017, pp. 1–12, Oct. 2017.

- [40] V. Choudhary, M. K. Dutta, and A. Singh, "Reversible watermarking scheme for authentication and integrity control in biometric images," in *Proc. 4th Int. Conf. Inf. Syst. Comput. Netw. (ISCON)*, Mathura, India, Nov. 2019, pp. 662–666.
- [41] S. Masmoudi, M. Charfeddine, and C. Ben Amar, "A semi-fragile digital audio watermarking scheme for MP3-encoded signals using Huffman data," *Circuits, Syst., Signal Process.*, vol. 39, no. 6, pp. 3019–3034, Jun. 2020.
- [42] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision—ECCV 2006 (Lecture Notes in Computer Science)*, vol. 3951, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Germany: Springer, 2006, pp. 404–417, doi: [10.1007/11744023_32](https://doi.org/10.1007/11744023_32).
- [43] J. A. Bloom, I. J. Cox, T. Kalker, J. P. M. G. Linnartz, M. L. Miller, and C. B. S. Traw, "Copy protection for DVD video," *Proc. IEEE*, vol. 87, no. 7, pp. 1267–1276, Jul. 1999.
- [44] M. Maes, T. Kalker, J.-P. M. G. Linnartz, J. Talstra, F. G. Depovere, and J. Haitsma, "Digital watermarking for DVD video copy protection," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 47–57, Sep. 2000.
- [45] R. Petrovic and V. Atti, "Watermark based access control to copyrighted content," in *Proc. 11th Int. Conf. Telecommun. Modern Satell., Cable Broadcast. Services (TELSIKS)*, Oct. 2013, pp. 315–322.
- [46] B. Abd-El-Atty, A. M. Ilyasu, H. Alaskar, A. El-Latif, and A. Ahmed, "A robust quasi-quantum walks-based steganography protocol for secure transmission of images on cloud-based E-healthcare platforms," *Sensors*, vol. 20, no. 11, p. 3108, 2020.
- [47] G. Zhou and D. Lv, "An overview of digital watermarking in image forensics," in *Proc. 4th Int. Joint Conf. Comput. Sci. Optim.*, Yunnan, PR, USA, Apr. 2011, pp. 332–335.
- [48] A. F. Qasim, R. Aspin, F. Meziane, and P. Hogg, "ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images," *Multimedia Tools Appl.*, vol. 78, no. 12, pp. 16433–16463, Jun. 2019.
- [49] L. Liu and X. Li, "Watermarking protocol for broadcast monitoring," in *Proc. Int. Conf. E-Bus. E-Government*, May 2010, pp. 1634–1637.
- [50] M. Parvaix, L. Girin, and J.-M. Brossier, "A watermarking-based method for informed source separation of audio signals with a single sensor," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1464–1475, Aug. 2010.
- [51] G. Bailly, V. Attina, C. Baras, P. Bas, S. Baudry, D. Beaudemps, R. Brun, J.-M. Chassery, F. Davoine, F. Elisei, G. Gibert, L. Girin, D. Grison, J.-P. Léoni, J. Liénard, N. Moreau, and P. Nguyen, "ARTUS: Synthesis and audiovisual watermarking of the movements of a virtual agent interpreting subtitling using Cued Speech for deaf televiewers," *Model., Meas. Control C*, vol. 67SH, Suppl. 2, pp. 177–187, 2006.
- [52] G. Tzanetakis, "Music information retrieval: Theory and applications," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 915–916.
- [53] E. Mezghani, M. Charfeddine, C. Ben Amar, and H. Nicolas, "Audiovisual video characterization using audio watermarking scheme," in *Proc. 15th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Dec. 2015, pp. 213–218.
- [54] M. Hirakawa and J. Iijima, "Mobile services and implementation of digital watermarks in audio files," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, vol. 31, Oct. 2010, pp. 94–97, doi: [10.1109/IIHMSP.2010.31](https://doi.org/10.1109/IIHMSP.2010.31).
- [55] M. O. Agbaje, O. Awodele, and A. C. Ogbona, "Big data, audience measurement and digital watermarking: A review," in *Proc. e-Skills Knowl. Prod. Innov. Conf.*, Cape Town, South Africa, 2014, pp. 17–28.
- [56] I. Portilla, "Television audience measurement: Proposals of the industry in the era of digitalization," *Trípodos*, no. 36, pp. 75–92, Jul. 2015.
- [57] F. Prior, M. L. Ingeholm, B. A. Levine, and L. Tarbox, "Potential impact of HITECH security regulations on medical imaging," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Piscataway, NJ, USA, Sep. 2009, pp. 60–2157.
- [58] P. Ruotsalainen, "Privacy and security in teleradiology," *Eur. J. Radiol.*, vol. 73, no. 1, pp. 31–35, Jan. 2010.
- [59] Y. Alkhourayyif, "National ID cards," *Int. J. Comput. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 44–48, 2013.
- [60] S. Hakak, A. Kamsin, O. Tayan, M. Y. I. Idris, A. Gani, and S. Zerdoumi, "Preserving content integrity of digital holy Quran: Survey and open challenges," *IEEE Access*, vol. 5, pp. 7305–7325, 2017.
- [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," in *Proc. Taibah Univ. Int. Conf. Adv. Inf. Technol. Holy Quran Its Sci.*, Madinah, Saudi Arabia, Dec. 2013, pp. 30–36.
- [62] N. S. Kamaruddin, A. Kamsin, and S. Hakak, "Associated diacritical watermarking approach to protect sensitive arabic digital texts," in *Proc. AIP Conf.*, Oct. 2017, Art. no. 020074.
- [63] H.-T. Hu and T.-T. Lee, "Hybrid blind audio watermarking for proprietary protection, tamper proofing, and self-recovery," *IEEE Access*, vol. 7, pp. 180395–180408, 2019, doi: [10.1109/ACCESS.2019.2958095](https://doi.org/10.1109/ACCESS.2019.2958095).
- [64] Q. Wu and M. Wu, "A novel robust audio watermarking algorithm by modifying the average amplitude in transform domain," *Appl. Sci.*, vol. 8, no. 5, p. 723, May 2018.
- [65] W. Lanxun, Y. Chao, and P. Jiao, "An audio watermark embedding algorithm based on mean-quantization in wavelet domain," in *Proc. 8th Int. Conf. Electron. Meas. Instrum.*, Xi'an, China, Aug. 2007, pp. 423–425.
- [66] Q. Wu and M. Wu, "Adaptive and blind audio watermarking algorithm based on chaotic encryption in hybrid domain," *Symmetry*, vol. 10, no. 7, p. 284, Jul. 2018, doi: [10.3390/sym10070284](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, Syst., Signal Process.*, vol. 35, no. 2, pp. 553–572, 2016.
- [68] J. Zhang and B. Han, "Robust audio watermarking algorithm based on moving average and DCT," 2017, *arXiv: 1704.02755*.
- [69] D. Pan, "A tutorial on MPEG/audio compression," *IEEE Multimedia-Mag.*, vol. 2, no. 2, pp. 60–74, Jun. 1995, doi: [10.1109/93.388209](https://doi.org/10.1109/93.388209).
- [70] J. D. Markel and A. H. Gray, "Linear Prediction of Speech," *J. Sound Vib.*, vol. 51, no. 4, p. 595, 1976.
- [71] J. Dittmann and C. Kraetzer, *Audio Benchmarking Tools and Steganalysis*, Standard IST-2002-507932, Revision, ECRYPT, European Network of Excellence in Cryptology, Network of Excellence, Information Society Technologies, 2006, p. 10-1.1, vol. 1.
- [72] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," *Proc. SPIE*, vol. 3657, pp. 226–239, Apr. 1999.
- [73] M. Kutter and F. Hartung, "Introduction to watermarking techniques," in *Information Hiding: Techniques for Steganographie and Digital Watermarking*, F. A. P. Petitcolas and S. Katzenbeisser Eds., 1st ed. Norwood, MA, USA: Artech House, 2000, pp. 97–120.
- [74] W.-S. Gan and S.-M. Kuo, *Embedded Signal Processing With the Micro Signal Architecture*. Hoboken, NJ, USA: Wiley, 2007.
- [75] X. Xiaojuan, H. Peng, and C. He, "DWT-based audio watermarking using support vector regression and subsampling," in *Proc. 7th Int. Workshop Fuzzy Log. Appl.*, 2007, pp. 136–144.
- [76] M. El'Arbi, M. Charfeddine, S. Masmoudi, M. Koubaa, and C. Ben Amar, "Video watermarking algorithm with BCH error correcting codes hidden in audio channel," in *Proc. IEEE Symp. Comput. Intell. (SSCI)*, Paris, France, Oct. 2011, pp. 17–164.
- [77] H.-T. Hu, L.-Y. Hsu, and H.-H. Chou, "Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity," *Digit. Signal Process.*, vol. 31, pp. 115–123, Aug. 2008.
- [78] N. Christian and H. Jurgen, "Digital watermarking and its influence on audio quality," in *Proc. 105th AES convention*, 1998, pp. 1–16.
- [79] A. G. Acevedo, "Audio watermarking quality evaluation," in *e-Business and Telecommunication Networks*, J. Ascenso, Ed. Cham, Switzerland: Springer, 2006, pp. 272–283.
- [80] W. Lu, Z. Chen, L. Li, X. Cao, J. Wei, N. Xiong, J. Li, and J. Dang, "Watermarking based on compressive sensing for digital speech detection and recovery," *Sensors*, vol. 18, no. 7, p. 2390, Jul. 2018, doi: [10.3390/s18072390](https://doi.org/10.3390/s18072390).
- [81] V. Bhat, G. I. Sen, and A. Das, "Audio watermarking based on quantization in wavelet domain," in *Proc. Int. Conf. Inf. Syst. Secur.*, 2008, pp. 235–242.
- [82] R. Martinez-Noriega, H. Kang, B. Kurkoski, K. Yamaguchi, K. Kobayashi, and M. Nakano, "Increasing robustness of audio watermarking DM using ATHC codes," in *Proc. Mexican Conf. Inform. Secur.*, Oaxaca, Mexico, 2006, pp. 1–4.
- [83] A.-H. Ali and M. Ahmad, "Digital audio watermarking based on the discrete wavelets transform and singular value decomposition," *Eur. J. Sci. Res.*, vol. 39, no. 1, pp. 6–21, 2010.
- [84] M. Şehirli, F. Gürgen, and S. İkizoğlu, "Performance evaluation of digital audio watermarking techniques designed in time, frequency and cepstrum domains," in *Proc. Int. Conf. Adv. Inf. Syst.*, Izmir, Turkey, vol. 3261, 2004, pp. 430–440.
- [85] U. Uludag and L.-M. Arslan, "Audio watermarking using DC level shifting," in *Advanced Topics in Speech Processing Project Report*, Turkey: Bogazici Univ., Electr. Electron. Eng. Department, Istanbul, 2001, pp. 1–6.
- [86] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, nos. 3–4, pp. 313–336, 1996.
- [87] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Process.*, vol. 66, no. 3, pp. 337–355, May 1998.

- [88] K. Khaldi and A.-O. Boudraa, "Audio watermarking via EMD," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 3, pp. 675–680, Mar. 2013.
- [89] P. K. Dhar and T. Shimamura, "A blind LWT-based audio watermarking using fast Walsh Hadamard transform and singular value decomposition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 125–128.
- [90] A. A. Mohammed, "Robust audio watermarking comparison between modified phase and wavelet based techniques," *Int. J. Comput. Sci. Eng.*, vol. 4, no. 6, pp. 1–7, 2015.
- [91] W. Diao, Y. Wu, W. Zhang, B. Liu, and N. Yu, "Robust audio watermarking algorithm based on air channel characteristics," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Guangzhou, China, Jun. 2018, pp. 288–293.
- [92] S. Esmaili, "Content based audio watermarking and retrieval using time frequency analysis," M.S. thesis, Dept. Elect. Comput. Eng., Ryerson Univ., Toronto, ON, Canada, 2004, doi: 10.32920/ryerson.14655894.v1.
- [93] K. Umapathy, B. Ghoraani, and S. Krishnan, "Audio signal processing using time-frequency approaches: Coding, classification, fingerprinting, and watermarking," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–28, Dec. 2010, doi: 10.1155/2010/451695.
- [94] M. A. Nematollahi, S. A. R. Al-Haddad, S. Doraisamy, and H. Gamboa-Rosales, "Speaker frame selection for digital speech watermarking," *Nat. Acad. Sci. Lett.*, vol. 39, no. 3, pp. 197–201, Jun. 2016, doi: 10.1007/s40009-016-0430-8.
- [95] A. Oermann, A. Lang, and C. Vielhauer, "Digital speech watermarking and its impact to biometric speech authentication," in *New Advances in Multimedia Security, Biometrics, Watermarking and Cultural Aspects*, J. Dittmann, C. Vielhauer, and J. Hansen, Eds. Berlin, Germany: Logos Verlag, 2006, pp. 33–51.
- [96] Y. Wang, Z. Liu, and J.-C. Huang, "Multimedia content analysis-using both audio and visual clues," *IEEE Signal Process. Mag.*, vol. 17, no. 6, pp. 12–36, Nov. 2000.
- [97] J. Vavrek, E. Vozarikova, M. Pleva, and J. Juhar, "Broadcast news audio classification using SVM binary trees," in *Proc. 35th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2012, pp. 469–473.
- [98] E. Mezghani, M. Charfeddine, and C. Ben Amar, "Audio silence deletion before and after MPEG video compression," in *Proc. Int. Conf. Comput. Appl. Technol.*, Sousse, Tunisia, 2013, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/ICCATT.2013.6521969>.
- [99] G. Sell and P. Clark, "Music tonality features for speech/music discrimination," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2014, pp. 2489–2493, doi: 10.1109/ICASSP.2014.6854048.
- [100] J. Becker, C. Rohlfing, "A segmental spectral flatness measure for harmonic-percussive discrimination," *Inst. Commun. Eng.*, RWTH Aachen Univ., Aachen, Germany, 2013.
- [101] H. Hermansky, N. Morgan, A. Bayya, and P. Kohn, "RASTA-PLP speech analysis technique," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dec. 1992, pp. 121–124.
- [102] B. K. Baniya and J. Lee, "Importance of audio feature reduction in automatic music genre classification," *Multimedia Tools Appl.*, vol. 75, no. 6, pp. 3013–3026, 2016.
- [103] G. Tzanetakis and F. Cook, "Sound analysis using MPEG compressed audio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Jun. 2000, pp. 11761–1176.
- [104] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Apr. 1997, pp. 1331–1334.
- [105] L. White and S. King. (2003). *The Eustace Speech Corpus*. Centre for Speech Technology Research, University of Edinburgh. [Online]. Available: <https://www.cstr.ed.ac.uk/projects/eustace>
- [106] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendmeier, and B. Weiss, "A database of German emotional speech," *Interspeech*, vol. 5, pp. 1517–1520, Sep. 2005.
- [107] B. K. Khonglah and S. R. Mahadeva Prasanna, "Speech/music classification using speech-specific features," *Digit. Signal Process.*, vol. 48, pp. 71–83, Jan. 2016.
- [108] C. Panagiotakis and G. Tziritas, "A speech/music discriminator based on RMS and zero-crossings," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 155–166, Feb. 2005.
- [109] K. El-Maleh, M. Klein, G. Petrucci, and P. Kabal, "Speech/music discrimination for multimedia applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, Jun. 2000, pp. 2445–2448.
- [110] Z.-H. Fu and J.-F. Wang, "Robust features for effective speech and music discrimination," in *Proc. ROCLING*, 2008, pp. 1–7.
- [111] M.-E. JE, S. Garcia-Galan, N. Ruiz-Reyes, P. Vera-Candeas, and F. Rivas-Peña, "Speech/music discrimination using a single warped lpc-based feature," in *Proc. ISMIR*, vol. 5, 2005, pp. 16–25.
- [112] S. Gaikwad, B. Gawali, and S. Mehrotra, "Gender identification using SVM with combination of MFCC," *Adv. Comput. Res.*, vol. 4, no. 1, pp. 69–73, 2012.
- [113] Y. Hu, D. Wu, and A. Nucci, "A Pitch-based gender identification with two-stage classification," *Secur. Commun. Netw.*, vol. 5, no. 2, pp. 211–225, 2012.
- [114] J. Ahmad, M. Fiaz, S.-I. Kwon, M. Sodanil, B. Vo, and S. W. Baik, "Gender identification using MFCC for telephone applications—A comparative study," 2016, *arXiv: 1601.01577*.
- [115] K. Dai, H. J. Fell, and J. MacAuslan, "Recognizing emotion in speech using neural networks," *Telehealth Assistive Technol.*, vol. 31, pp. 38–43, Apr. 2008.
- [116] S. Haq, P. J. B. Jackson, and J. Edge, "Audio-visual feature selection and reduction for emotion classification," in *Proc. Int. Conf. Auditory-Visual Speech Process.*, Tantalooma, QLD, Australia, 2008, pp. 1–6.
- [117] F. Shah, "Discrete wavelet transforms and artificial neural networks for speech emotion recognition," *Int. J. Comput. Theory Eng.*, vol. 2, no. 3, p. 319, 2010.
- [118] M. M. Javidi and E. F. Roshan, "Speech emotion recognition by using combinations of C5.0, neural network (NN), and support vector machines (SVM) classification methods," *J. Math. Comput. Sci.*, vol. 6, no. 3, pp. 191–200, Apr. 2013.
- [119] M. Sheikhan, M. Bejani, and D. Gharavian, "Modular neural-SVM scheme for speech emotion recognition using ANOVA feature selection method," *Neural Comput. Appl.*, vol. 23, no. 1, pp. 22–215, 2013.



MAHA CHARFEDDINE (Member, IEEE) was born in March 1981. She received the B.S. degree in computer science engineering and the M.S. and Ph.D. degrees in computer science from the National Engineering School of Sfax (ENIS), Tunisia, in 2005, 2007, and 2013, respectively. She is currently an Assistant Professor with the Computer Sciences and Applied Mathematics Department, 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 of the Research Group with the Intelligent Machines of REGIM-Laboratory (LR11ES48). She co-supervises graduate students in master's and Ph.D. degrees. Her research interests include 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 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 received the B.S. degree in telecommunication engineering from the National Engineering School of Tunis (ENIT), in 2009, and the M.S. and Ph.D. degrees in computer science from the National Engineering School of Sfax, Tunisia, in 2012 and 2018, respectively. She is also a Research Member with the Research Group of the Intelligent Machine REGIM-Laboratory, ENIS. Her research interests include audio analysis and classification, artificial intelligence, digital watermarking, and signal processing.



SALMA MASMOUDI (Member, IEEE) was born in Sfax, Tunisia, in 1984. She received the B.S. degree in computer science engineering and the M.S. degree in computer science from the National Engineering School of Sfax (ENIS), Tunisia, in 2008 and 2010, respectively. She is currently pursuing the Ph.D. degree in computing systems engineering with the Laboratory (REGIM-Lab), National School of Engineering of Sfax, University of Sfax, Tunisia. She worked as a Contractual

Assistant with the National School of Engineering of Sfax, between 2011 and 2014. Her research interests include digital audio watermarking and MP3 compression.



CHOKRI BEN AMAR (Senior Member, IEEE) received the B.S. degree in electrical engineering from the National Engineering School of Sfax (ENIS), in 1989, and the M.S. and Ph.D. degrees in computer engineering from the National Institute of Applied Sciences of Lyon, France, in 1990 and 1994, respectively.

He spent one year with the University of Haute Savoie, France, as a Teaching Assistant and a Researcher before joining the Higher School of Sciences and Techniques of Tunis (ESSTT) as an Assistant Professor, in 1995. In 1999, he joined Sfax University (USS) as an Assistant Professor, and since 2011, he has been a Full Professor with the Department of Computer Sciences and Applied Mathematics, National Engineering School of Sfax. Since September 2018, he has been a Full Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests include computer vision, image and video analysis, intelligent algorithms and their applications to data classification and approximation, pattern recognition, watermarking, image and video indexing, and securing. He founded the IEEE Signal Processing Society (SPS) Tunisia Chapter, in January 2009, and is actually the Chair of this Chapter.

During this period, the chapter was organized five IEEE Distinguished Lectures and other technical and professional activities. He has been an Advisor of the IEEE SPS Student Chapter at 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 several 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.

...