Digital Signal Processing 120 (2022) 103251

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Efficient watermarking algorithm for digital audio/speech signal

M. Yamni [a,*], H. Karmouni [a], M. Sayyouri [b], H. Qjidaa [a]

[a] CED-ST, STIC, Laboratory of Electronic Signals and Systems of Information LESSI, Faculty of Science Dhar El Mahrez, University Sidi Mohamed Ben Abdellah, Fez, Morocco

[b] Engineering, Systems and Applications Laboratory, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, BP 72, My Abdallah Avenue Km. 5 Imouzzer Road, Fez, Morocco

A R T I C L E I N F O

Article history:
Available online 24 September 2021

Keywords:
Audio/speech watermarking
Embedding algorithm
Discrete Tchebichef moment transform
Discrete wavelet transform
Chaotic system

A B S T R A C T

Ensuring conflicting requirements such as imperceptibility, payload capacity, and robustness, becomes a great challenge for a robust audio/speech watermarking algorithm. To achieve this, this paper proposes a blind and robust audio/speech watermarking algorithm that combines the discrete Tchebichef moment transform (DTMT), the chaotic system of the mixed linear–nonlinear coupled map lattices (MLNCML), and discrete wavelet transform (DWT). The watermark is encrypted by the MLNCML system, and then it is embedded in the norm of the Tchebichef moments of the low frequency components. DTMT and MLNCML are used to achieve high robustness, high payload capacity, and high security, while DWT is used to achieve a satisfactory level of imperceptibility. In addition, the adopted strategy has a blind nature, where no original audio/speech is needed in watermark extraction. Compared to other existing audio/speech watermarking algorithms, the proposed algorithm gives better results in terms of robustness and payload capacity while keeping the embedding effect non-perceptible and undetectable.

1. Introduction

Audio/speech watermarking has been widely used to solve some problems related to copyright protection, monitoring, indication of content manipulation, fingerprinting, and information carrier [1], [2].

Several requirements have been stated by the International Federation of the Phonographic Industry (IFPI) that must be insured to achieve an efficient audio watermarking. Indeed, audio/speech watermarking algorithms should meet [2] 1) the imperceptibility requirement, where algorithms must embed the watermark without affecting the perceptual quality of the host audio/speech signal with a SNR greater than 20 dB for watermarked audio/speech; 2) the robustness requirement, where common signal processing attacks do not prevent watermark detection; 3) the payload capacity requirement, where the algorithms must be able to embed an amount of data (watermark) greater than 20 bps (bits per second) into the host signal; and 4) the security requirement, where the watermark can only be detected by the authorized person.

A variety of researches have been published that aim to meet these basic requirements. Works have been published where watermarking is implemented in the time domain [3], [4] and other works have been published where watermarking is implemented in transform domains such as such discrete wavelet transform (DWT) [5], [6], [7], [8], [9], lifting wavelet transform (LWT) [10], [11], [12], discrete cosine transform (DCT) [13], [14], [15], [9], singular value decomposition (SVD) [16], [5], [13], [10], [14], [17], cepstrum [18], [19], [20], etc. Transform-domain methods have attracted considerable attention due to their efficiency in terms of robustness [15].

To achieve an imperceptible and undetectable watermark in the host's audio signal, Erçelebi and Batakçı [11] proposed an audio watermarking scheme in the lifting-based wavelet domain, where the watermark is embedded in the low frequency components. Bhat et al. [16] proposed a robust and blind audio watermarking algorithm based on SVD and quantization index modulation (QIM). The watermark insertion and extraction procedures are based on quantization of the norm of singular values of the blocks. The same authors proposed in [5] a secure, robust, and blind adaptive audio watermarking algorithm based on SVD in the DWT domain using synchronization code. The watermark is embedded by applying a QIM process on the singular values in the SVD of the wavelet domain blocks. Based on SVD–DCT with chaotic synchronization code technique, Lei et al. [13] proposed a blind and robust audio watermarking scheme where the watermark is blindly embedded into

* Corresponding author.
E-mail addresses: mohamed.yamni@usmba.ac.ma (M. Yamni), hicham.karmouni@usmba.ac.ma (H. Karmouni), mhamed.sayyouri@usmba.ac.ma (M. Sayyouri), qjidah@yahoo.fr (H. Qjidaa).

https://doi.org/10.1016/j.dsp.2021.103251
1051-2004/

the high-frequency band of the SVD–DCT block. To reduce processing times, the same authors [10] combined LWT with QIM method and SVD and constructed a robust and blind audio watermarking scheme for copyright protection. Yong-mei et al. [14] proposed a blind audio/speech watermarking scheme based on DWT and SVD for copyright protection. Despite the inaudible and robust nature of this schema, it has the drawback that it is not robust against random cropping and time-scale modification. Wang et al. [6] proposed a blind and adaptive audio watermarking algorithm based on vector norm and DWT, where the watermark bits were embedded in the vector norm of the segmented approximation components after DWT. Though it showed good imperceptibility and high capacity (up to 102.4 bps), its robustness against amplitude scaling was not very good. Hu et al. [7] introduced a variable-dimensional vector modulation (VDVM) scheme to maximize the efficiency of the norm-space DWT-based blind audio watermarking technique, which allows to achieve a high capacity (up to 301.46 bps) while guaranteeing imperceptivity and robustness. A blind audio watermarking method based on LWT and QRD is proposed by Dhar et al. [12] for audio copyright protection, where the watermark is embedded into the largest element of the upper triangular matrix obtained from the low frequency LWT coefficients of each original audio frame. Hu and Hsu [15] proposed a robust, transparent and high capacity blind audio watermarking by exploiting perceptual masking in the DCT domain. A spread spectrum (SS)-based audio watermarking technique which involves the psychoacoustic model, multiple scrambling, adaptive synchronization, frequency alignment, and coded-image watermark is proposed in [20]. Hu and al. [8] proposed an effective blind speech watermarking via adaptive mean modulation (AMM) and package synchronization in DWT domain. The watermark is embedded into the low frequency coefficients in DWT and a synchronization code is embedded into the high frequency coefficients in DWT as a location indicator. Based on SVD using Angle-Quantization (AQ), Ogura et al. [17] proposed a blind audio watermarking scheme, where the watermark data is embedded into the angle between the largest singular value and the second largest singular value of each diagonal matrix by quantization. This scheme has the advantage of a high payload capacity (up to 172.39 bps). Liu et al. [21] defined and discussed the feature coefficients cross-correlation degree (CCCD) of speech signal, and then explored a watermark embedding method based on the feature, in order to enlarge the embedding capacity and solve the security issue of watermark schemes based on public features. Based on localized features, Kaur et al. [22] proposed an adaptive audio watermarking algorithm where the capacity of this algorithm is adaptive and dependent on the nature of the watermarked audio signal. Saadi et al. [9] proposed a blind watermarking algorithm of speech and audio signals based on DWT and DCT transformations and based on sub-sampling as decomposition technique for correlation purpose. The watermark is embedded in the vector norm of DCT domain, which improve the robustness of the algorithm. This scheme allows to obtain a relatively high imperceptibility and a good robustness against various attacks except the robustness against the MP3 compression (64 kbps) which was not very good, moreover the capacity of this algorithm is not too high (up to 57 bps).

The aforementioned watermarking methods have advantages and disadvantages. The main weaknesses of these algorithms are the low payload capacity and the low robustness against some common signal processing attacks particularly cropping and MP3 compression.

Over the last years, discrete Tchebichef moment transform (DTMT also known as discrete Tchebichef moments) due to its ability to represent global features has found extensive applications in image processing field such as image watermarking [23], [24], [25], [26], image reconstruction [27], [28], [29], image analysis [27], pattern recognition [30], texture retrieval [31], image compression [32], focus measuring [33], medical image analysis [34], and image projection [35]. A detailed description of discrete orthogonal moment transforms is given in [36],[37],[38],[39],[40],[41].

The DTMT has the characteristic of being natively orthogonal and discrete, which has established it as high discriminative power transform. Furthermore, DTMT does not involve any numerical approximation in the computation process [27], and has a lower computational complexity and it does not require complex transform unlike continuous transforms. The use of DTMT produces minimal reconstruction error in applications that require the reconstruction procedure such as image watermarking and image reconstruction [24], [29]. Other useful properties that make this transform desirable are discussed in [29], such as signal decorrelation and energy compaction. These properties can be exploited to build an efficient audio/speech watermarking algorithm.

This paper proposes an efficient method of watermarking audio/speech signals. For this purpose, the discrete Tchebichef moment transform (DTMT), the chaotic system of the mixed linear–nonlinear coupled map lattices (MLNCML) [42], and the discrete wavelet transform (DWT) are combined for the first time.

The core idea is to embed an encrypted watermark in the norm of the Tchebichef moment vectors of the low frequency signal coefficients. The watermark is embedded into the host audio/speech signal by three main steps. First, the original digital audio/speech is segmented, and then the three-level DWT decomposition of each segment is applied to obtain the low frequency coefficients. Second, for the low frequency coefficients of each segment, DTMT is used to generate a set of discrete orthogonal vectors of Tchebichef moments, in which the watermark will be embedded in the norm of the vectors of Tchebichef moments. Finally, inverse DTMT and DWT are performed to obtain the watermarked audio/speech signal.

In order to eliminate the spatial relationship between the bits of the watermark and improve the robustness of the audio/speech watermarking, the chaotic system of MLNCML [42] is employed to encrypt the watermark before embedding it in the original audio/speech signal. The MLNCML system also ensures the security of the proposed watermarking algorithm.

The combination of DTMT, MLNCML, and DWT allows to obtain high-performance audio/speech watermarking in terms of robustness, payload capacity, and security while keeping imperceptibility at a satisfactory level. In addition, the adopted strategy has a blind nature, where no original audio/speech is needed in hidden watermark extraction. Experimental results indicate that the watermarked audio/speech is perceptually similar to the original audio/speech (SNR is greater than 29 dB and 27 dB for audio and speech, respectively), even though the payload capacity of the proposed algorithm is high (between 469.1 bps and 801.8 bps). In addition, they indicate that the hidden watermark is very robust against wide class of common signal processing attacks such as resampling, noise corruption, requantization, echo addition, random cropping, low-pass filtering, high-pass filtering, amplitude scaling and MP3 compression. Another advantage is that our algorithm offers low processing time and easier hardware implementation. The results are presented and compared with existing audio/speech watermarking algorithms to verify the effectiveness of our algorithm.

The rest of the paper is organized as follows. In Section 2, some preliminaries about DWT, DTMT, and MLNCML are provided. Section 3 presents the proposed audio/speech watermarking algorithm. Experiments to test its performance are carried out in Section 4. Finally, conclusion is drawn in Section 5.

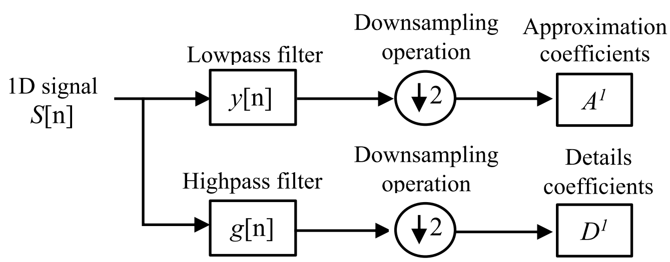


Fig. 1. One-level DWT decomposition.

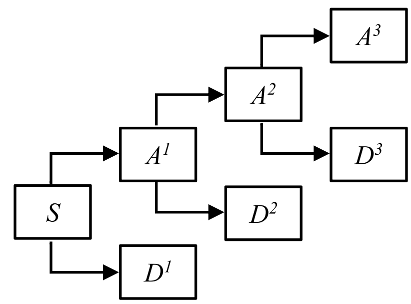


Fig. 2. 3-level DWT decomposition.

2. Preliminaries

DWT, DTMT, and MLNCML play an important role in the proposed watermarking method, so they will be discussed in detail in this section.

2.1. Discrete Wavelet Transform (DWT)

The DWT is a transform, widely used in signal processing, that can give a time-frequency representation of a digital signal [43]. DWT can be applied on one-dimensional signals (audio or speech signal) and on two-dimensional signals (image or video frames). DWT is based on high pass and low pass filtering and sampling operations applied using a tree-based recursive algorithm. For one-dimensional signals, the DWT produces two sets of coefficients [44] (Fig. 1): the approximation coefficients A^1 (low frequencies) which are produced by passing the signal S through a low-pass filter y, and the detail coefficients D^1 (high frequencies) which are produced by passing the signal S through a high-pass filter g.

Depending on the purpose, the low frequencies part A^1 might be further decomposed into two parts of low frequencies A^2 and high frequencies D^2 [9]. Likewise, A^2 can also be broken down into two parts of low frequencies A^3 and high frequencies D^3, etc. Fig. 2 shows a 3-level DWT decomposition of signal S. The inverse DWT (iDWT) makes it possible to reconstruct the signal S from its wavelet subbands.

Several existing orthogonal wavelet bases namely *Haar* wavelets, *Daubechies* wavelets, *Coiflets* wavelets and *Symlets* wavelets, can be used. In this paper, we apply the DWT decomposition of signal S using the most popular wavelet base, the *Haar* wavelet, implemented in MATLAB to obtain approximation coefficients, which are less sensitive to the human hearing system. The use of these coefficients, when embedding the watermark into the host signal, makes the watermark strong and inaudible in our watermarking algorithm. The smaller level DWT will influence the robustness of the watermark; and the larger one will cause large calculation, so 3-level DWT is performed in our watermarking algorithm.

2.2. Discrete Tchebichef moment transform (DTMT)

The 1-D DTMT is defined by projecting a signal function onto a set of orthogonal discrete Tchebichef polynomials. The DTMT of order n is defined as [29]

$$M_n = \sum_{x=0}^{N-1} \frac{t_n(x;N)}{\sqrt{\rho(n,N)}} f(x), \quad n = 0,1,2,......N-1 \tag{1}$$

where $f(x)$ is an audio/speech signal of length N; $t_n(x;N)$ is the nth order discrete Tchebichef polynomial, defined as [45]

$$t_n(x;N) = (1-N)_n \sum_{k=0}^{n} \frac{(-n)_k(-x)_k(1+n)_k}{(k!)^2(1-N)_k}, \quad n,x,=0,1,2,....,N-1 \tag{2}$$

with $(x)_k = x(x+1)(x+2)....(x+k-1)$ is the Pochhammer symbol and $\rho(n,N)$ is the norm of Tchebichef polynomial, defined as

$$\rho(n,N) = (2n)! \binom{N+n}{2n+1} \tag{3}$$

The discrete Tchebichef polynomials satisfy the following orthogonality relation [24]

$$\sum_{x=0}^{N-1} t_n(x;N)t_m(x;N) = \rho(n,N)\delta_{n,m}, \quad n,m,=0,1,2,....,N-1 \tag{4}$$

and a three-term recurrence relation of type [46]

$$\begin{aligned}&(n+1)t_{n+1}(x;N) - (2n+1)(2x-N+1)t_n(x;N)\\ &\quad + n(N^2-n^2)t_{n-1}(x;N) = 0\\ &n = 1,2,....,N-1; \quad x = 0,1,....,N-1\end{aligned} \tag{5}$$

with the initial conditions

$$t_0(x;N) = 1 \quad \text{and} \quad t_1(x;N) = 2x-N+1 \tag{6}$$

Eq. (4) leads to the inverse DTMT (iDTMT), which is given as follows

$$\tilde{f}(x) = \sum_{n=0}^{N-1} M_n \frac{t_n(x;N)}{\sqrt{\rho(n,N)}}, \quad x = 0,1,2,......N-1 \tag{7}$$

where N is the maximum order of moments and $\tilde{f}(x)$ is the reconstructed audio/speech signal.

Using DTMT produces minimal reconstruction error because it is orthogonal and exactly satisfies the orthogonality property in discrete domain and does not require any numerical approximation [29]. In addition, this transform is discrete, like the digital audio/speech signal, so there is no discretization error.

In the proposed watermarking algorithm, the watermark is embedded in the DTMT domain after selecting the low frequency coefficients of audio/speech signal.

2.3. Mixed linear–nonlinear coupled map lattice (MLNCML)

The logistic map was originally proposed by May [47]. It is a first-order difference equation represented by $\tau(x) = \mu x(1-x)$. Zhang et al. in [42] proposed the chaotic system of the mixed linear–nonlinear coupled map lattices (MLNCML), which considers *Len* logistic maps coupled by neighborhood links and Arnold cat map links as follows:

$$\begin{aligned}x_{n+1}(i) = {}&(1-\varepsilon)\tau[x_n(i)] + (1-\eta)\frac{\varepsilon}{2}\{\tau[x_n(i+1)]\\ &+\tau[x_n(i-1)]\} + \eta\frac{\varepsilon}{2}\{\tau[x_n(j)]+\tau[x_n(k)]\}\end{aligned} \tag{8}$$

where $i, j, k (1 \le i, j, k \le Len)$ are the lattices; $\varepsilon (0 \le \varepsilon \le 1)$ and $\eta (0 \le \eta \le 1)$ are the coupling parameters; n is the time index $(n = 1, 2, 3, \ldots)$, and $\tau(x)$ is logistic mapping defined as $\tau(x) = \mu x(1-x)$, $\mu \in (0, 4]$. The relations of i, j, k are defined by the Arnold cat map described in following equation:

$$\begin{bmatrix} j \\ k \end{bmatrix} = \begin{bmatrix} 1 & p \\ q & pq+1 \end{bmatrix} \begin{bmatrix} i \\ i \end{bmatrix} (\mathrm{mod}\, L) \tag{9}$$

where p and q are the parameters of Arnold cat map, and $\mathrm{mod}\, L$ is modulo L (Euclidian division rest).

In order to eliminate the spatial relationship between the bits of the watermark and improve the robustness of the proposed audio/speech watermarking algorithm, the MLNCML is employed to encrypt the watermark before embedding it in the original audio/speech signal.

Based on this chaotic system, a $N \times M$ chaotic matrix $H = \{h(i, j), 0 \le i < N, 0 \le j < M\}$ is generated, and then a binary chaotic matrix $H_b = \{h_b(i, j), 0 \le i < N, 0 \le j < M\}$ is obtained by binarizing H as follows:

$$h_B(i, j) = \begin{cases} 1, & \text{if } h(i, j) \ge e \\ 0, & \text{if } h(i, j) < e \end{cases}, \quad (0 \le i < N, 0 \le j < M) \tag{10}$$

the threshold e is the mean of H.

The binary chaotic matrix H_b will be used to encrypt the watermark (Section 3.1) before embedding it in the host audio/speech signal.

Due to the initial sensitivity of the chaotic system, the initial value of the MLNCML (including ε, η, μ and the initial value x_0 of logistic mapping) can be considered as a secret key to ensuring the security of the proposed algorithm.

3. The proposed algorithm

The main idea of the proposed algorithm is to segment the host audio/speech signal into many sections, and then, for each segment, the 3-level DWT decomposition is applied to obtain low frequency coefficients. After that, for the low frequency coefficients of each segment, the DTMT is used to generate a set of discrete orthogonal vectors of Tchebichef moments, in which a watermark, encrypted by MLNCML, will be embedded in the norm of Tchebichef moment vectors. Finally, iDTMT and iDWT are performed to obtain the watermarked audio/speech signal. Segmenting and cutting techniques are adopted to improve the robustness and reduce the computational complexity of the proposed algorithm, respectively. The embedding model is shown in Fig. 3.

In watermark extraction, the watermark can be extracted from the watermarked audio/speech signal without using the original audio/speech signal, which shows the blind nature of the proposed algorithm.

Let $S = \{s(i), 0 \le i < length\}$ denote a host digital audio/speech signal with *length* samples, and $W = \{w(i, j), 0 \le i < N, 0 \le j < M\}$ denotes a $N \times M$ binary image to be embedded within the host audio/speech signal, and $w(i, j) \in \{0, 1\}$ is the bit value at position (i, j). The subsections below explain more the processes of the proposed watermarking algorithm.

3.1. Watermark preprocessing

To improve the robustness and enhance the security of the algorithm, the watermark is first encrypted from W to $W_1 = \{w_1(i, j), 0 \le i < N, 0 \le j < M\}$ by applying an XOR operation between the binary chaotic matrix H_b (Section 2.3) and the watermark W, that is

$$W_1 = XOR(H_b, W) \tag{11}$$

Then, it is transformed into a one-dimensional sequence of ones and zeros as follows:

$$W_2 = \{w_2(k), 0 \le k < L_W, L_W = N \times M\} \tag{12}$$

The initial values of the MLNCML (including ε, η, μ and the initial value x_0 of logistic mapping) are used as security keys during the watermark embedding process. These values are noted "Key1". Therefore, it is impossible to extract the watermark from the watermark signal without this secret key.

3.2. Watermark embedding

(1) Dividing the encrypted watermark sequence W_2 into L_1 sections, where each section $PW_2(m)$ having L_{seg} bits, as follows

$$\begin{aligned} & L_{seg} = L_W / L_1 \\ & PW_2(m) = \{pw_2(m)(n) = w_2(n + m \times L_{seg}), \\ & 0 \le m < L_1, 0 \le n < L_{seg}\} \end{aligned} \tag{13}$$

L_1 is a constant and is chosen to be 4 in our experiment.

(2) In order to improve the robustness of proposed scheme against cropping, segment the input audio/speech signal S into L_{seg} audio/speech segments, where each segment $PS(n)$ having L samples, as follows

$$PS(n) = \{ps(n)(i) = s(i + n \times L),\ 0 \le n < L_{seg},\ 0 \le i < L\} \tag{14}$$

(3) For each audio/speech segment $PS(n)$, 3-level DWT is performed to obtain the wavelet coefficients of $A^3(n)$, $D^3(n)$, $D^2(n)$, $D^1(n)$, where $A^3(n)$ is the segment of low frequency coefficients and $D^1(n), D^2(n), D^3(n)$ are the detail segment signals.

To take the advantage of low frequency coefficients, which has a higher energy value and robustness against various signal processing, the low frequency coefficients $A^3(n)$ are only considered in the following steps.

(4) In order to increase the payload capacity of the proposed algorithm and to reduce the computational complexity in the DTMT calculation (next step), the $A^3(n)$ is cut into L_1 sections, and each section $PA^3(n)$ having L_2 coefficients, where

$$\begin{aligned} & L_2 = \frac{L}{L_1 \times 2^3} \\ & PA^3(n)(m) = \{pa^3(n)(m)(j) \\ & = a^3(n)(j + m \times L_2), \quad 0 \le n < L_{seg}, \quad 0 \le m < L_1, \quad 0 \le j < L_2\} \end{aligned} \tag{15}$$

(5) Calculating the Tchebichef moments of $PA^3(n)(m)$, that is

$$M(n)(m) = PA^3(n)(m).T^t \tag{16}$$

where $M(n)(m)$ and $PA^3(n)(m)$ are $1 \times L_2$ vectors, and T is the $L_2 \times L_2$ matrix polynomials of Tchebichef

$$T = \left\{ \frac{t_n(x; L_2)}{\sqrt{\rho(n, L_2)}}, 0 \le n, x < L_2 \right\} \tag{17}$$

here, T^t denotes the transpose of the matrix T,

The matrix form in Eq. (16) facilitates the implementation of Tchebichef moments and accelerates the calculation process.

(6) The vector of Tchebichef moments $M(n)(m)$ is decomposed into two (correlated) moment sub-vectors $M_1(n)(m)$ and $M_2(n)(m)$ using the following sub-sampling operations:

$$\begin{matrix} M_1(n)(m)(k) = M(n)(m)(2k) \\ M_2(n)(m)(k) = M(n)(m)(2k-1) \end{matrix}, \quad k = 1, 2, \ldots, L_2/2 \tag{18}$$

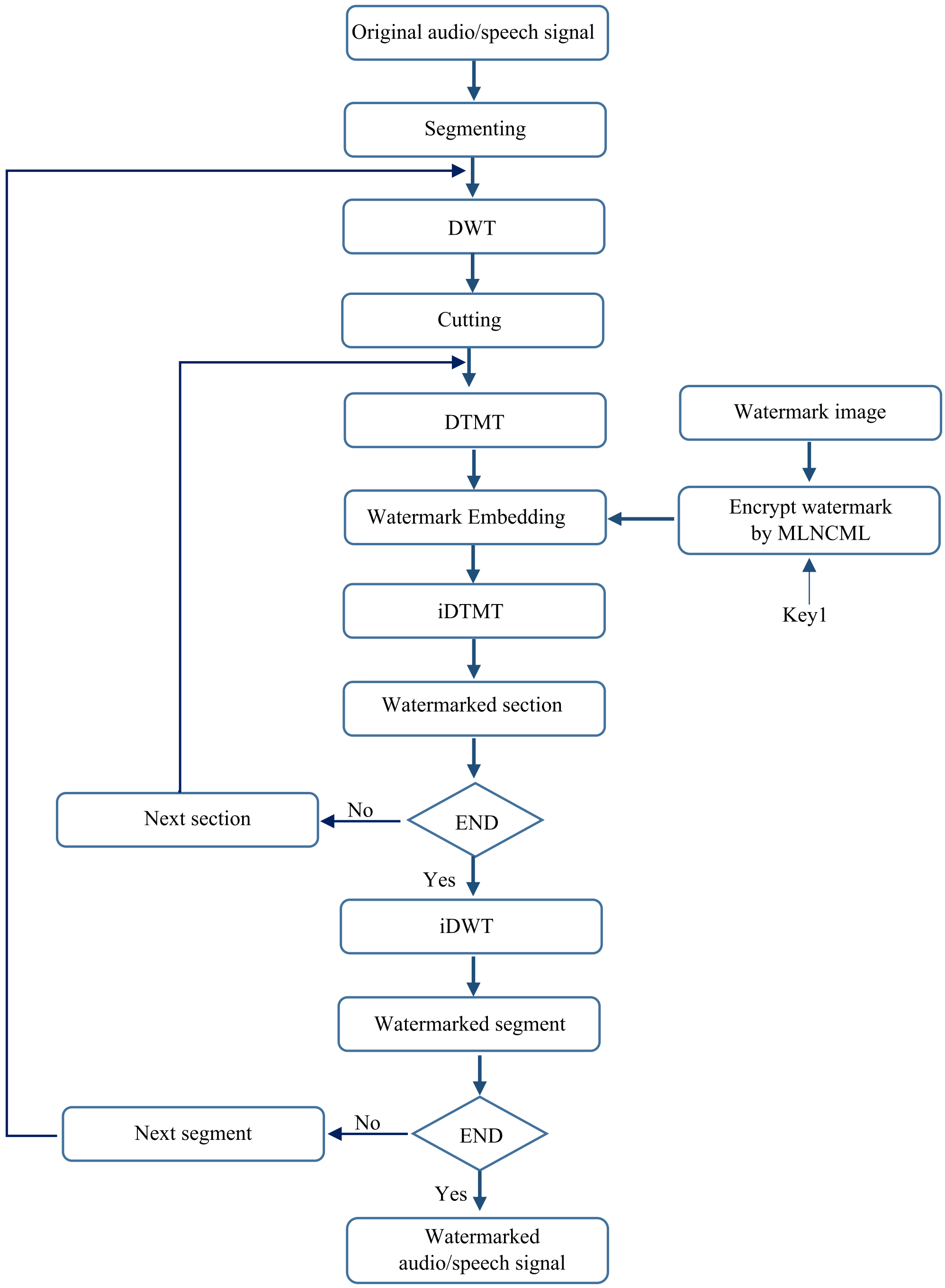


Fig. 3. Watermark embedding.

(7) Calculating σ_1 and σ_2 the norm of $M_1(n)(m)$ and $M_2(n)(m)$, respectively, that is

$$\sigma_1 = \sqrt{\sum_{k=0}^{L_2/2} M_1(n)(m)(k)^2}, \quad \sigma_2 = \sqrt{\sum_{k=0}^{L_2/2} M_2(n)(m)(k)^2} \tag{19}$$

(8) In order to guarantee the high robustness and a satisfactory transparency of the watermark, the proposed algorithm embeds a watermark bit in the norm space of Tchebichef moment coefficients. The embedding rule is given as follows:

$$\overline{M_1(n)(m)} = \begin{cases} \left(\frac{\sigma_1+\sigma_2}{2} + \Delta\right) \frac{M_1^t(n)(m)}{\sigma_1}, & \text{if } pw_2(m)(n) = 1 \\ \left(\frac{\sigma_1+\sigma_2}{2} - \Delta\right) \frac{M_1^t(n)(m)}{\sigma_1}, & \text{if } pw_2(m)(n) = 0 \end{cases} \tag{20}$$

$$\overline{M_2(n)(m)} = \begin{cases} \left(\frac{\sigma_1+\sigma_2}{2} - \Delta\right) \frac{M_2^t(n)(m)}{\sigma_2}, & \text{if } pw_2(m)(n) = 1 \\ \left(\frac{\sigma_1+\sigma_2}{2} + \Delta\right) \frac{M_2^t(n)(m)}{\sigma_2}, & \text{if } pw_2(m)(n) = 0 \end{cases} \tag{21}$$

where $0 \le n < L_{seg}, 0 \le m < L_1$, $\overline{M_1(n)(m)}$ and $\overline{M_2(n)(m)}$ are the watermarked vectors of Tchebichef moments, $M_1(n)(m)$ and $M_2(n)(m)$ are the original vectors of Tchebichef moments, Δ denotes the watermark embedding strength (also called the quantization step).

The quantization step Δ influences the imperceptibility and robustness of the proposed audio/speech watermarking algorithm. The smaller quantization step will influence the robustness of the watermark, and the larger one will cause low imperceptibility (SNR < 20 dB).

The imperceptibility must be ensured first, because if the quality of the audio/speech signal cannot be preserved method will not be accepted neither by industry nor users [2]. Therefore, the value of quantization step Δ should be as large as possible under the constraint of imperceptibility. Based on the observations made on dozens of different types of audio/speech signal, we obtained the optimal quantization $\Delta = 0.05$ which gives a SNR greater than 20 dB.

(9) Combine the two vectors $\overline{M_1(n)(m)}$ and $\overline{M_2(n)(m)}$ using the opposite operation in step (6) produce the watermarked vector of Tchebichef moments $\overline{M(n)(m)}$:

$$\begin{array}{l} \overline{M(n)(m)}(2k) = \overline{M_1(n)(m)}(k) \\ \overline{M(n)(m)}(2k-1) = \overline{M_2(n)(m)}(k) \end{array}, \quad k = 1, 2, ..., L_2/2 \tag{22}$$

(10) Calculating the iDTMT of $\overline{M(n)(m)}$ to obtain the watermarked section $\overline{PA^3(n)(m)}$

$$\overline{PA^3(n)(m)} = \overline{M(n)(m)}.T \tag{23}$$

(11) By substituting the $PA^3(n)(m)$ with $\overline{PA^3(n)(m)}$, we get the watermarked low frequency coefficients $\overline{A^3(n)}$, and then 3-level iDWT is performed in order to obtain the watermarked audio/speech segment $\overline{PS(n)}$.

(12) Get the watermarked audio/speech signal by combining the watermarked audio/speech segment signals.

All the variables of the proposed algorithm are explained in the experiments section.

3.3. Watermark extraction

The adopted strategy achieves a blind extraction; it is meant that no original audio/speech is needed in watermark extraction.

Let S^* denote the watermarked audio/speech signal. The watermark extraction procedure is summarized as follows.

(1) Split the watermarked audio/speech S^* into L_{seg} segments, and then the 3-level DWT is performed on each audio/speech segment $PS^*(n)$ to get the coefficients $A^{*3}(n), D^{*3}(n), D^{*2}(n)$ and $D^{*1}(n)$.

(2) The $A^{*3}(n)$ is cut into L_1 sections, and for each section $PA^{*3}(n)(m)$ apply steps ($5 \sim 7$) of the embedding process, in order to obtain the norms σ_1^* and σ_2^*.

(3) Extract the bit sequence $PW_2^*(m)$ from $PA^{*3}(n)(m)$ by using the following extraction rule:

$$pw_2^*(m)(n) = \begin{cases} 1, \text{ if } \sigma_1^* > \sigma_2^* \\ 0, \text{ if } \sigma_1^* \le \sigma_2^* \end{cases}, \quad 0 \le m < L_1, \quad 0 \le n < L_{seg} \tag{24}$$

where $PW_2^*(m) = \{pw_2^*(m)(n), 0 \le m < L_1, 0 \le n < L_{seg}\}$

(4) All sequences PW_2^* are combined to get the watermark sequence $W_2^* = \{w_2^*(k), 0 \le k < L_W, L_W = N \times M\}$, and then W_2^* is reorganized into a two-dimensional image W_1^* of size $N \times M$ bits.

(5) The secret key "Key1" (same initial values of MLNCML used in embedding) is used to generate a binary chaotic matrix $H_b^* = \{h_b^*(i, j), 0 \le i \le N, 0 \le j \le M\}$.

(6) Finally, the extracted watermark W^* can be obtained by decrypting W_1^* as follows

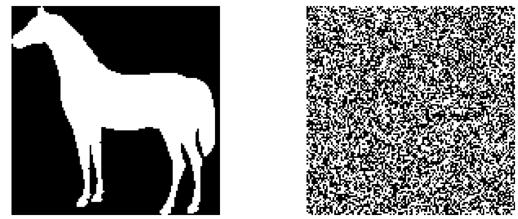
$$W^* = XOR(H_b^*, W_1^*) \tag{25}$$

4. Experimental results

This section presents experimental results that evaluate the performance of the proposed audio/speech watermarking algorithm. The simulations were conducted in a Matlab R2018a environment on different audio/speech signals (Table 1) which are selected from [48]. Each audio/speech is a 16-bit mono signal in the WAVE format sampled at $F_s = 44100$ Hz. A $N \times M$ binary image (where $N, M = 128$) is used as watermark to embed in the audio/speech signal, shown in Fig. 4-(a). The binary watermark encrypted by MLNCML system is shown in Fig. 4-(b). We also tested many other kinds of audio/speech with different watermarks, and the simulation results are similar.

Table 1
Audio/speech information.

Audio/Speech	Type	Contents	Duration (s)
35_Glockenspiel_1	Audio	Glockenspiel	35
44_Soprano_1	Audio	Soprano	28
47_bass_1	Audio	Bass	30
49_spfe_1	Speech	Female speech (English)	23
51_spff_1	Speech	Female speech (French)	21
52_spmf_1	Speech	Male speech (French)	24



(a) (b)

Fig. 4. Watermark image: (a) the original watermark, and (b) the encrypted watermark.

The *Haar* wavelet basis is used with three decomposition levels. The parameters used to start the proposed algorithm are: $L_w = N \times M = 128 \times 128 = 16384$, $L_1 = 4$, $L_{seg} = L_w/L_1 = 4096$, the duration of the audio/speech signal d (is given in Table 1), $F_s = 44100$ Hz, $L = floor(d \times F_s/L_{seg})$, $L_2 = L/(L_1 \times 2^3)$, and the quantization step is chosen to be $\Delta = 0.05$. All these parameters have been chosen so as to achieve a good compromise between the contending requirements of payload capacity, imperceptibility, and robustness.

In order to evaluate the efficiency of the proposed algorithm, several comparisons were made in terms of payload capacity, imperceptibility and robustness with other existing audio and speech watermarking algorithms [9], [6], [7], [12], [17], [5], [20], [8], [14], [21], which were named in abbreviated form as the DWT-DCT-norm [9], DWT-norm [6], DWT-VDVM-norm [7], LWT-QRD [12], SVD-AQ [17], DWT-SVD-QIM [5], SS [20], DWT-AMM [8], DWT-SVD [14], CCCD [21].

4.1. Payload capacity test

Payload capacity (Eq. (26)) is defined as the number of bits that are embedded into the audio/speech signal within a unit of time, denoted by P and measured in the unit of bps (bit per-second).

$$P = \frac{B}{d}(\text{bps}) \tag{26}$$

where, B is the number of embedded bits in the original audio/speech signal, and d is the duration of the audio/speech in seconds.

Table 2 shows the payload capacity of our algorithm, for all selected audio/speech signals, and shows that the payload capacity of our proposed algorithm is very high where it is between 469.1 bps and 801.8 bps. Therefore, the proposed algorithm will be very desirable for the copyright protection of audio/speech signals and for hiding secret messages in the host audio/speech signals, hence it can be easily modified for steganography purposes.

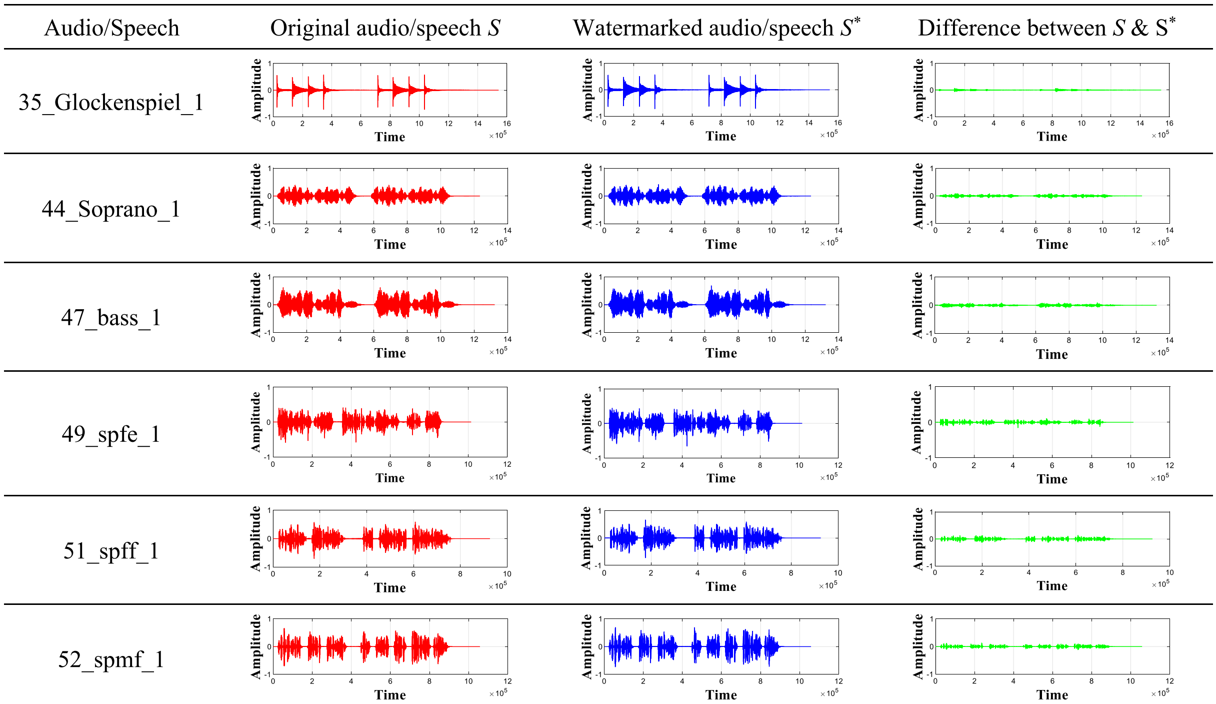


Fig. 5. Original and watermarked audio/speech signals and the difference between them by using the proposed algorithm.

Table 2
Payload capacity of the proposed algorithm (in bps) for different audio and speech signals.

Audio/Speech	Type	Duration (s)	Payload capacity (bps)
35_Glockenspiel_1	Audio	35	469.14
44_Soprano_1	Audio	28	595.94
47_bass_1	Audio	30	558.22
49_spfe_1	Speech	23	722.95
51_spff_1	Speech	21	801.81
52_spmf_1	Speech	24	700

Table 3
SNR and ODG for different audio/speech signals using the proposed algorithm.

Type	Audio/Speech	SNR(dB)	ODG
Audio	35_Glockenspiel_1	30.4009	−0.2320
	44_Soprano_1	28.46	−0.2514
	47_bass_1	28.5500	−0.2483
Speech	49_spfe_1	27.0470	−0.2639
	51_spff_1	26.9009	−0.2706
	52_spmf_1	27.5029	−0.2602

Table 4
Comparison with SNR, ODG and payload capacity for audio signals.

Watermarking algorithm	Average SNR (dB)	Average ODG	Average payload capacity (bps)
Proposed	29.1370	−0.2439	541.100
DWT-norm [6]	22.46	−0.53	102.40
DWT-DCT-norm [9]	31.0786	–	40.2700
DWT-VDVM-norm [7]	20.400	−0.151	301.46
DWT-SVD-QIM [5]	24.37	–	45.9
SVD-AQ [17]	30.3	–	172.39
DWT-SVD [14]	21.2	–	27.56
SS [20]	30.1	–	17.2

4.2. Imperceptibility test

The imperceptibility test aims to measure the perceptual quality or perceptual transparency of the embedded watermarking in the original audio/speech signals. The quality of the watermarked audio/speech signal was evaluated objectively using signal-to-noise ratio (SNR) [6], defined as follows:

$$\mathrm{SNR}(S, S^*) = 10\log_{10}\left(\frac{\sum_{i=0}^{length-1} s^2(i)}{\sum_{i=0}^{length-1}[s(i)-s^*(i)]^2}\right) \tag{27}$$

where S and S^* are the original and the watermarked audio/speech signals, respectively.

The recommendation of the IFPI [49], [10], states that the SNR must be greater than 20 dB to have an imperceptible watermarked audio/speech signal.

We also use the perceptual evaluation of audio quality (PEAQ) [50] to simulate the subjective evaluations made by human subjects. The PEAQ renders an objective difference grade (ODG) between −4 and 0, signifying a perceptual impression from "very annoying" to "imperceptible". In this study, the PEAQ metric used for the imperceptibility test was a Matlab implementation [51] released by the TSP Lab at McGill University [50]. We adopted the basic version of this software tool to assess the perceived quality of watermarked audio/speech signal.

Fig. 5 shows the original and the corresponding watermarked audio/speech and the differences between them respectively, which show the imperceptibility of the proposed algorithm. We can see that there is hardly any difference between the original audio/speech signals and the watermarked ones, which indicates that our algorithm has a good imperceptibility.

Table 3 presents the ODG's and SNR's of different watermarked audio/speech, where the ODG's are very close to zero and the SNR's are greater than 20 dB, satisfying the IFPI requirement. Therefore, the embedded watermark is virtually transparent. These results have illustrated the inaudible nature of our watermarking algorithm.

Tables 4 and 5 present the average SNR's and ODG's along with the payload capacities for the compared watermarking algorithms.

Table 5
Comparison with SNR, ODG and payload capacity for speech signals.

Watermarking algorithm	Average SNR (dB)	Average ODG	Average payload capacity (bps)
Proposed	27.1503	−0.2649	741.5867
DWT-DCT-norm [9]	30.1833	–	54.8267
LWT-QRD [12]	34.58	−0.73	172.39
DWT-AMM [8]	21.932	–	200
DWT-SVD [14]	20.7	–	27.56
CCCD [21]	25.77	−0.6520	49

ODG values are replaced by the "-" symbol for algorithms that do not specify the ODG value. We can see that the payload capacity of our proposed algorithm is much higher than that of other algorithms. Indeed, the average payload capacity of our algorithm is equal to 541.100 bps and 741.5867 bps for audio and speech signals, respectively, which shows the superiority of our watermarking algorithm in terms of payload capacity. It should be noted that the watermark embedding in the proposed algorithm does not affect the quality of the audio signal, because the average SNR is above 20 dB and the average ODG is very close to zero. Besides, the imperceptibility of the proposed algorithm is relatively lower than that of DWT-DCT-norm [9], SVD-AQ [17], SS [20], and LWT-QRD [12], but it is relatively high compared to the other selected algorithms such as the DWT-norm [6], DWT-VDVM-norm [7], DWT-SVD-QIM [5], DWT-AMM [8], DWT-SVD [14], and CCCD [21].

4.3. Robustness test

Algorithm robustness is defined as the ability of the algorithm to extract the embedded watermark after intentional or unintentional signal processing attacks. The robustness of our algorithm is evaluated using the Bit Error Rate (BER) defined in Eq. (28) [52] along with the Normalized Correlation (NC) defined in Eq. (29) [52].

$$\mathrm{BER}(W, W^*) = \frac{B_{ERR}}{B_T} \times 100\% \tag{28}$$

$$\mathrm{NC}(W, W^*) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} w(i,j)w^*(i,j)}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} w^2(i,j)}\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} w^{*2}(i,j)}} \tag{29}$$

where $W = \{w(i,j), 0 \le i < N, 0 \le j < M\}$ and $W^* = \{w^*(i,j), 0 \le i < N, 0 \le j < M\}$ denote the original and extracted watermark image, respectively. B_{ERR} is the number of erroneously extracted bits and B_T is the total number of watermark bits.

BER = 0 and NC = 1 mean that the attack doesn't have any effect on the watermark and the extraction is successful.

In order to illustrate the robust nature of our watermarking algorithm, attacks resampling, noise corruption, requantization, echo addition, random cropping, low-pass filtering, high-pass filtering, amplitude scaling and MP3 compression are used to estimate the robustness of our algorithm. Table 6 provides details of each type of attack.

It should be noted that the influence of external disturbances, modeling errors, and various uncertainties which are common in practical applications is not addressed in this paper because this influence is mainly related to the reconstruction of an unknown signal from a degraded signal and not to the watermarking of audio and speech signals. For more details, see [53], [54], [55], [56], [57].

Tables 7 and 8 present the robustness results of the proposed algorithm obtained against the attacks presented in Table 6 for different audio/speech signals. It is clearly shown that most of the attacks such as A, B, C, D, E, F, G, H, M, N do not lead to any extraction error where BER = 0 and NC = 1. We also see that the BERs are very low and the NCs are very high for the other attacks (I, J, K, L), where the maximum BER is about 8.84% for audios and 10.55% for speechs and the minimum NC is about 0.87 for audios and about 0.84 for speechs. Fig. 6, which shows some attacked watermarked signals ("35_Glockenspiel_1" and "49_spfe_1") and their extracted watermarks, confirms the above discussion. These experimental results clearly show the high robustness of the proposed watermarking algorithm against different attacks for different audio/speech signals.

In Tables 9 and 10, we summarize the proposed watermark extraction results comparing with that of the DWT-DCT-norm [9], DWT-norm [6] and LWT-QRD [12] against various attacks. These algorithms were chosen because they presented a higher performance in terms of robustness than other existing audio/speech watermarking algorithms, such as [13], [10], [11], [16], [5], as indicated in [9], [6] and [12]. The results of Tables 9 and 10 clearly show the superiority of the proposed algorithm in terms of robustness against various attacks compared to the algorithms [9], [6] and [12].

Despite the success of the proposed algorithm, it also has a drawback. The proposed watermarking method and all the methods reported in this section are not suitable for sensitive signals such as medical signals. Indeed, the strategies adopted consist in embedding a watermark in the original signal under the constraint of imperceptibility (the algorithm should offer more than 20 dB SNR for watermarked signal versus original signal). However, medical signals have a fairly high degree of sensitivity (especially regions of interest) where any modification of their contents negatively influences the exploitation and interpretation of these signals (even the algorithm offers more than 20 dB SNR for the watermarked signal compared to the original signal).

In [58], [59], [26], we have proposed robust algorithms called zero-watermarking algorithms to protect the copyright of color and grayscale medical images and color stereo images. The proposed zero-watermarking algorithms do not embed any information into the original images; instead, they construct a zero watermark using the image features, which can effectively protect the copyright of images while ensuring their integrity, they are very suitable for this type of images.

In order to solve the imperceptibility problem for medical signals, we aim in future work to build a robust zero watermarking algorithm for medical signals by combining the transforms adopted in this paper.

4.4. Estimation of computational complexity

In this section, the complexity of the proposed algorithm is estimated and analyzed in terms of speed. Figs. 7 and 8 show the time processing in the proposed watermarking algorithm for audio "35_Glockenspiel_1" and for speech "49_spfe_1", respectively. These figures show that for watermark embedding, our watermarking algorithm takes about 26.44 seconds and 24.93 seconds for audio "35_Glockenspiel_1" and speech "49_spfe_1", respectively, and for watermark extraction, it takes about 14.76 seconds and 13.36 seconds for audio "35_Glockenspiel_1" and speech "49_spfe_1", respectively. The most time consuming steps are the calculation of DWT and iDWT. It should be noted that DTMT and iDTMT take a very low computation time because we have implemented them based on matrix forms (Eq. (16) and Eq. (23)) and we have adopted the cutting technique (step 4 in Sections 3.2 and 3.3), which reduces the total complexity of our watermarking algorithm.

These numerical values are obtained using Matlab R2018a on a PC Intel(R) Core(TM) i3 CPU @ 2.40 GHz with 6 GB of RAM.

Table 6
Information on attack types and specifications.

Item	Attack Type	Description
A	Resampling	Conducting down-sampling to 22050 Hz and then up-sampling back to 44100 Hz.
B	Noise corruption (I)	Adding zero-mean white Gaussian noise to the watermarked audio/speech signal with SNR = 30 dB.
C	Noise corruption(II)	Adding zero-mean white Gaussian noise to the watermarked audio/speech signal with SNR = 20 dB.
D	Requantization	Quantizing the watermarked signal to 8 bits/sample and then back to 16bits/sample.
E	Echo addition(I)	Adding an echo signal with a delay of 50 ms and a decay to 5% to the watermarked audio/speech signal.
F	Echo addition(II)	Adding an echo signal with a delay of 300 ms and a decay to 40% to the watermarked audio/speech signal.
G	Cropping(I)	20% of the watermarked signal samples were randomly set to zero.
H	Cropping (II)	30% of the watermarked signal samples were randomly set to zero.
I	Cropping(III)	40% of the watermarked signal samples were randomly set to zero.
J	Low-pass filtering(I)	Applying a low-pass filter with a cutoff frequency of 4 kHz.
K	Low-pass filtering(II)	Applying a low-pass filter with a cutoff frequency of 500 Hz.
L	High-pass filtering	Applying a high-pass filter with a cutoff frequency of 200 Hz.
M	Amplitude scaling(I)	Scaling the amplitude of the watermarked signal by 0.7.
N	Amplitude scaling(II)	Scaling the amplitude of the watermarked signal by 1.3.
O	MP3 compression(I)	Applying MPEG-1 Layer 3 compression with 128 kbps to the watermarked audio/speech signal.
P	MP3 compression(II)	Applying MPEG-1 Layer 3 compression with 64 kbps to the watermarked audio/speech signal.

Table 7
BER and NC for different audio signals.

	BER (in %)			NC		
Attack type	35_Glockenspiel_1	44_Soprano_1	47_bass_1	35_Glockenspiel_1	44_Soprano_1	47_bass_1
No attack	00	00	00	1	1	1
A	00	00	00	1	1	1
B	00	00	00	1	1	1
C	00	00	00	1	1	1
D	00	00	00	1	1	1
E	00	00	00	1	1	1
F	00	00	00	1	1	1
G	00	00	00	1	1	1
H	00	00	00	1	1	1
I	0.3906	0.4456	0.3967	0.9941	0.9933	0.9941
J	0.4869	0.5038	0.4395	0.9961	0.9924	0.9934
K	0.5646	0.6287	0.6104	0.9916	0.9906	0.9909
L	2.4475	3.2471	2.6878	0.9638	0.9518	0.9601
M	00	00	00	1	1	1
N	00	00	00	1	1	1
O	00	00	00	1	1	1
P	7.5989	8.8440	7.6416	0.8890	0.8711	0.8886

Table 8
BER and NC for different speech signals.

	BER (in %)			NC		
Attack type	49_spfe_1	51_spff_1	52_spmf_1	49_spfe_1	51_spff_1	52_spmf_1
No attack	00	00	00	1	1	1
A	00	00	00	1	1	1
B	00	00	00	1	1	1
C	00	00	00	1	1	1
D	00	00	00	1	1	1
E	00	00	00	1	1	1
F	00	00	00	1	1	1
G	00	00	00	1	1	1
H	00	00	00	1	1	1
I	1.5015	1.6235	1.3489	0.9776	0.9758	0.9799
J	0.5188	0.5371	0.5005	0.9922	0.9919	0.9925
K	1.2817	1.5808	1.0315	0.9809	0.9765	0.9846
L	3.4424	5.6641	2.8442	0.9487	0.9171	0.9579
M	00	00	00	1	1	1
N	00	00	00	1	1	1
O	00	00	00	1	1	1
P	9.1371	10.5530	9.9243	0.8679	0.8476	0.8556

5. Conclusion

In this paper, we proposed a blind and robust watermarking algorithm for audio and speech signals based on DTMT, MLNCML, and DWT. First, after segmenting the original audio/speech signal, the low frequency coefficients are selected from each segment using the 3-level DWT, which ensures the imperceptibility of the proposed algorithm. Second, each vector of the low frequency coefficients is cut into several sections, and then a watermark encrypted by MLNCML is embedded in the norm of the vectors of Tchebichef moments, which gives a high robustness, high payload capacity, and high security.

Experiments were carried out on different audio and speech signals of different lengths, indicating that even though the pay-

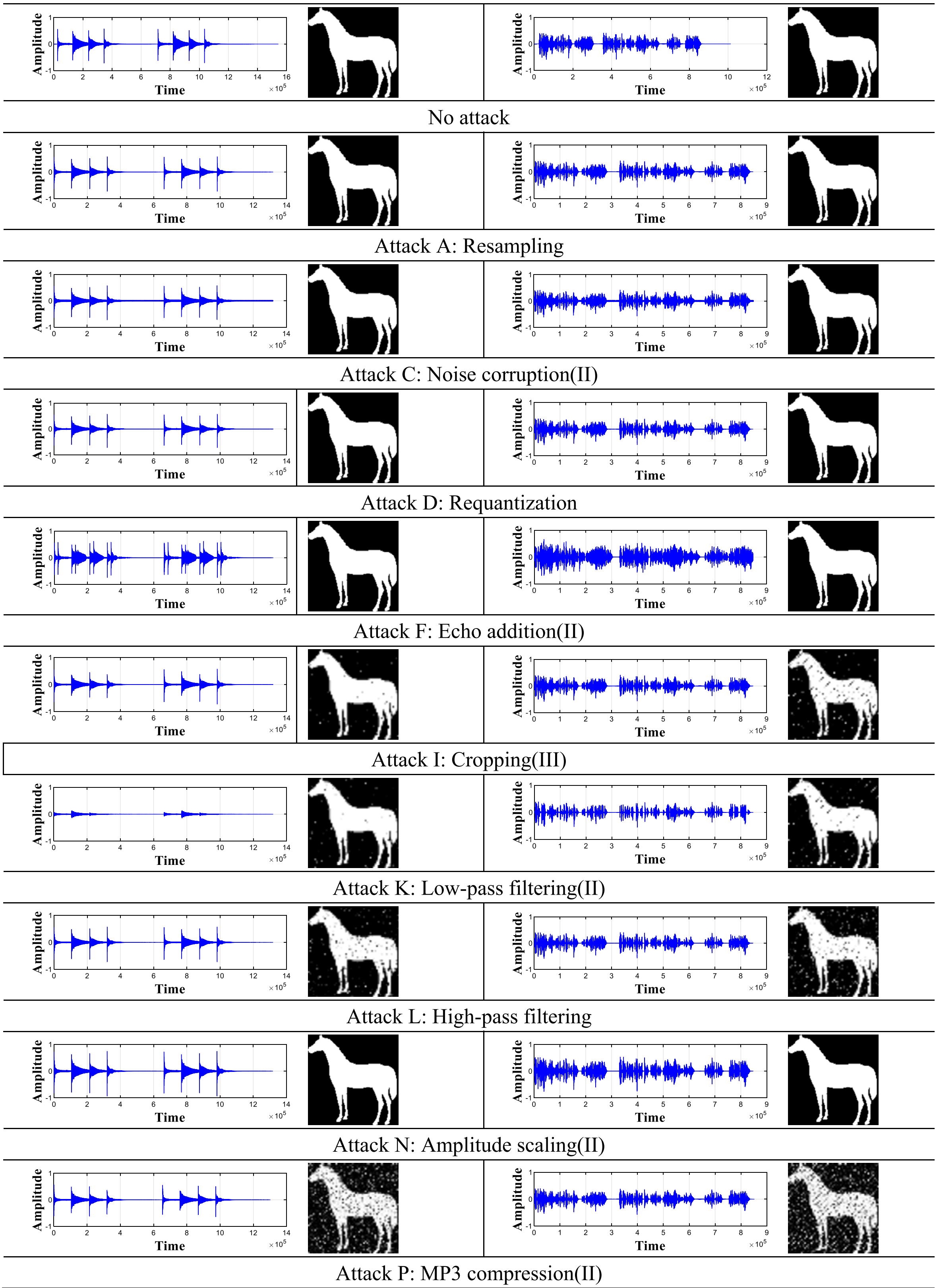


Fig. 6. The attacked watermarked signal and the extracted watermark: The left column shows results obtained from audio “35_Glockenspiel_1” while the right column shows results obtained from speech “49_spfe_1”.

load capacity of the proposed algorithm is very high, between 469.1 bps and 801.8 bps, the watermarked audio/speech is practically perceptually indistinguishable from the original audio/speech signal, because the embedding of all watermarks resulted in an average SNR ranging from 27.15 dB to 29.13 dB with an average ODG ranging from −0.264 to −0.243. Experiments also indicate that the watermark keeps good quality and robustness even though the watermarked audio/speech signal is attacked by vari-

Table 9
Robustness comparisons between the proposed algorithm and the algorithms DWT-DCT-norm [9] and DWT-norm [6] for the audio signals.

Attack type	Average BER (in %)			Average NC		
	Proposed	DWT-DCT-norm [9]	DWT-norm [6]	Proposed	DWT-DCT-norm [9]	DWT-norm [6]
No attack	00	00	00	1	1	1
A	00	5.6389	00	1	0.9520	1
B	00	00	00	1	1	1
C	00	00	00	1	1	1
D	00	00	00	1	1	1
E	00	00	00	1	1	1
F	00	00	7.9477	1	1	0.9307
G	00	00	0.5985	1	1	0.9937
H	00	00	0.9460	1	1	0.9859
I	0.4110	1.0742	14.8349	0.9938	0.9910	0.8723
J	0.4767	0.5865	0.8118	0.9940	0.9914	0.9878
K	0.6012	11.7657	1.8494	0.9910	0.8301	0.9721
L	2.7941	6.4629	3.5736	0.9586	0.9068	0.9501
M	00	00	27.7649	1	1	0.6579
N	00	00	25.1282	1	1	0.6933
O	00	00	00	1	1	1
P	8.0282	26.6907	00	0.8829	0.6745	1

Table 10
Robustness comparisons between the proposed algorithm and the algorithms DWT-DCT-norm [9] LWT-QRD [12] for the speech signals.

Attack type	Average BER (in %)			Average NC		
	Proposed	DWT-DCT-norm [9]	LWT-QRD [12]	Proposed	DWT-DCT-norm [9]	LWT-QRD [12]
No attack	00	00	00	1	1	1
A	00	5.2750	18.6646	1	0.9601	0.7673
B	00	00	00	1	1	1
C	00	00	00	1	1	1
D	00	00	00	1	1	1
E	00	00	14.1113	1	1	0.8190
F	00	00	38.1042	1	1	0.5442
G	00	00	8.5347	1	1	0.9306
H	00	00	14.4226	1	1	0.8134
I	1.4913	3.6499	23.2178	0.9778	0.9459	0.7182
J	0.5188	0.5447	3.0734	0.9922	0.9915	0.9428
K	1.2980	10.2396	21.2585	0.9807	0.8601	0.7363
L	3.9836	14.0320	31.3171	0.9412	0.8259	0.6154
M	00	00	00	1	1	1
N	00	00	1.5015	1	1	0.9776
O	00	00	3.32	1	1	0.9493
P	9.8715	25.3570	46.7224	0.8570	0.6995	0.4893

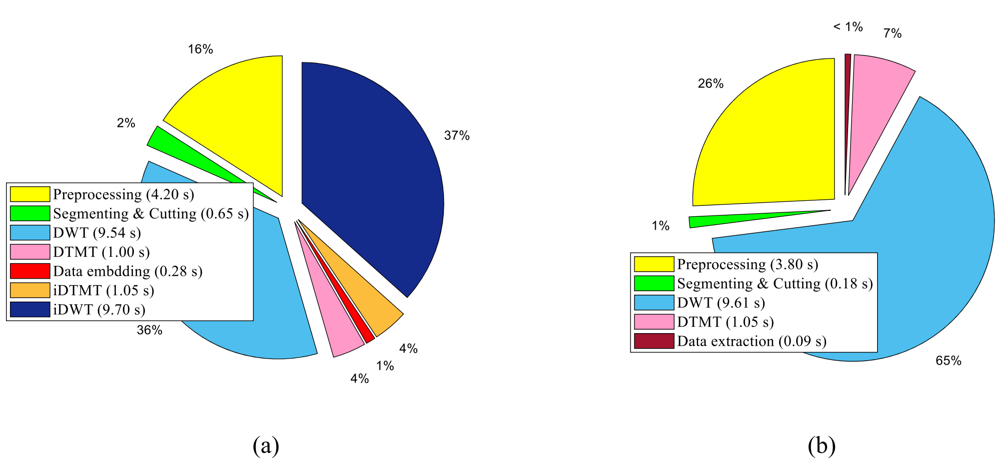
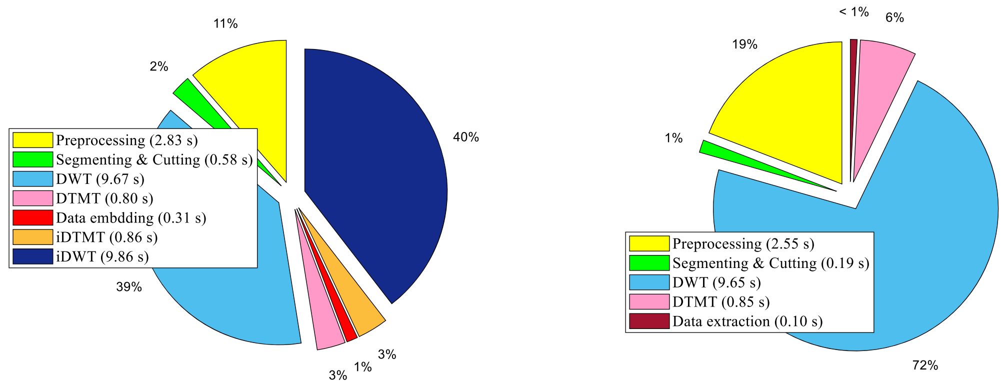


Fig. 7. Time processing in the proposed watermarking algorithm for audio "35_Glockenspiel_1" (a) the watermark embedding and (b) the watermark extraction. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

ous common signal processing attacks such as resampling, noise corruption, requantization, echo addition, random cropping, low-pass filtering, high-pass filtering, amplitude scaling and MP3 compression. In addition, the watermark can be extracted from the watermarked audio/speech signal without using the original audio/speech signal, which shows the blind nature of the proposed algorithm.

The performance of the proposed algorithm is superior to that of similar robust audio watermarking algorithms and speech watermarking algorithms.



(a) (b)

Fig. 8. Time processing in the proposed watermarking algorithm for speech “49_spfe_1” (a) the watermark embedding and (b) the watermark extraction.

Abbreviations

IFPI	International Federation of the Phonographic Industry
DWT	Discrete Wavelet Transform
iDWT	Inverse DWT
DTMT	Discrete Tchebichef Moment Transform
iDTMT	Inverse DTMT
MLNCML	Mixed Linear–Nonlinear Coupled Map Lattice
SNR	Signal-to-Noise Ratio
ODG	Objective Difference Grade
BER	Bit Error Rate
NC	Normalized Correlation

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

References

[1] M. Arnold, Audio watermarking: features, applications and algorithms, in: 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings, in: Latest Advances in the Fast Changing World of Multimedia, vol. 2, 2000, pp. 1013–1016 (Cat. No. 00TH8532).
[2] M.D. Swanson, B. Zhu, A.H. Tewfik, Current state of the art, challenges and future directions for audio watermarking, in: Proceedings IEEE International Conference on Multimedia Computing and Systems, vol. 1, 1999, pp. 19–24.
[3] P. Bassia, I. Pitas, N. Nikolaidis, Robust audio watermarking in the time domain, IEEE Trans. Multimed. 3 (2) (2001) 232–241.
[4] W.-N. Lie, L.-C. Chang, Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification, IEEE Trans. Multimed. 8 (1) (2006) 46–59.
[5] V. Bhat, I. Sengupta, A. Das, An adaptive audio watermarking based on the singular value decomposition in the wavelet domain, Digit. Signal Process. 20 (6) (2010) 1547–1558.
[6] X. Wang, P. Wang, P. Zhang, S. Xu, H. Yang, A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform, Signal Process. 93 (4) (2013) 913–922.
[7] H.-T. Hu, L.-Y. Hsu, H.-H. Chou, Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity, Digit. Signal Process. 31 (2014) 115–123.
[8] H.-T. Hu, S.-J. Lin, L.-Y. Hsu, Effective blind speech watermarking via adaptive mean modulation and package synchronization in DWT domain, EURASIP J. Audio Speech Music Process. 2017 (1) (2017) 10.
[9] S. Saadi, A. Merrad, A. Benziane, Novel secured scheme for blind audio/speech norm-space watermarking by Arnold algorithm, Signal Process. 154 (Jan. 2019) 74–86, https://doi.org/10.1016/j.sigpro.2018.08.011.
[10] B. Lei, Y. Soon, F. Zhou, Z. Li, H. Lei, A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition, Signal Process. 92 (9) (2012) 1985–2001.
[11] E. Erçelebi, L. Batakçı, Audio watermarking scheme based on embedding strategy in low frequency components with a binary image, Digit. Signal Process. 19 (2) (2009) 265–277.
[12] P.K. Dhar, A blind audio watermarking method based on lifting wavelet transform and QR decomposition, in: 8th International Conference on Electrical and Computer Engineering, 2014, pp. 136–139.
[13] B.Y. Lei, Y. Soon, Z. Li, Blind and robust audio watermarking scheme based on SVD–DCT, Signal Process. 91 (8) (2011) 1973–1984.
[14] Y. Cai, W. Guo, H. Ding, An audio blind watermarking scheme based on DWT-SVD, J. Softw. 8 (7) (July 2013) 1801–1808.
[15] H.-T. Hu, L.-Y. Hsu, Robust, transparent and high-capacity audio watermarking in DCT domain, Signal Process. 109 (2015) 226–235.
[16] V. Bhat, I. Sengupta, A. Das, A new audio watermarking scheme based on singular value decomposition and quantization, Circuits Syst. Signal Process. 30 (5) (2011) 915–927.
[17] M. Ogura, Y. Sugiura, T. Shimamura, SVD based audio watermarking using angle-quantization, in: 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2017, pp. 119–122.
[18] X. Li, H.H. Yu, Transparent and robust audio data hiding in cepstrum domain, in: 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings, in: Latest Advances in the Fast Changing World of Multimedia, vol. 1, 2000, pp. 397–400 (Cat. No. 00TH8532).
[19] H.-T. Hu, W.-H. Chen, A dual cepstrum-based watermarking scheme with self-synchronization, Signal Process. 92 (4) (Apr. 2012) 1109–1116, https://doi.org/10.1016/j.sigpro.2011.11.001.
[20] Y. Lin, W.H. Abdulla, Audio Watermark, vol. 146, Springer, 2015.
[21] Z. Liu, J. Huang, X. Sun, C. Qi, A security watermark scheme used for digital speech forensics, Multimed. Tools Appl. 76 (7) (2017) 9297–9317.
[22] A. Kaur, M.K. Dutta, K.M. Soni, N. Taneja, Localized & self adaptive audio watermarking algorithm in the wavelet domain, J. Inf. Secur. Appl. 33 (2017) 1–15.
[23] L. Zhang, G. Qian, W. Xiao, Z. Ji, Geometric invariant blind image watermarking by invariant Tchebichef moments, Opt. Express 15 (5) (2007) 2251–2261.
[24] C. Deng, X. Gao, X. Li, D. Tao, A local Tchebichef moments-based robust image watermarking, Signal Process. 89 (8) (2009) 1531–1539.
[25] F. Ernawan, M.N. Kabir, An improved watermarking technique for copyright protection based on Tchebichef moments, IEEE Access 7 (2019) 151985–152003.
[26] M. Yamni, H. Karmouni, M. Sayyouri, H. Qjidaa, J. Flusser, Novel Octonion Moments for color stereo image analysis, in: Digital Signal Processing, Oct. 2020, p. 102878.
[27] R. Mukundan, S.H. Ong, P.A. Lee, Image analysis by Tchebichef moments, IEEE Trans. Image Process. 10 (9) (2001) 1357–1364.
[28] R. Mukundan, Improving image reconstruction accuracy using discrete orthonormal moments, 2003.
[29] H. Zhu, M. Liu, H. Shu, H. Zhang, L. Luo, General form for obtaining discrete orthogonal moments, IET Image Process. 4 (5) (2010) 335–352.
[30] H. Zhu, H. Shu, T. Xia, L. Luo, J.L. Coatrieux, Translation and scale invariants of Tchebichef moments, Pattern Recognit. 40 (9) (2007) 2530–2542.
[31] J.V. Marcos, G. Cristóbal, Texture classification using discrete Tchebichef moments, JOSA A 30 (8) (2013) 1580–1591.
[32] N.A. Abu, S.L. Wong, N.S. Herman, R. Mukundan, An efficient compact Tchebichef moment for image compression, in: 10th International Conference on Information Science, Signal Processing and Their Applications (ISSPA 2010), 2010, pp. 448–451.

[33] P.T. Yap, P. Raveendran, Image focus measure based on Chebyshev moments, IEE Proc., Vis. Image Signal Process. 151 (2) (2004) 128–136.
[34] X.B. Dai, H.Z. Shu, L.M. Luo, G.-N. Han, J.-L. Coatrieux, Reconstruction of tomographic images from limited range projections using discrete Radon transform and Tchebichef moments, Pattern Recognit. 43 (3) (2010) 1152–1164.
[35] N.A. Abu, W.S. Lang, S. Sahib, Image projection over the edge, in: 2010 Second International Conference on Computer and Network Technology, 2010, pp. 344–348.
[36] O. El ogri, A. Daoui, M. Yamni, H. Karmouni, M. Sayyouri, H. Qjidaa, 2D and 3D medical image analysis by discrete orthogonal moments, Proc. Comput. Sci. 148 (2019) 428–437.
[37] H. Karmouni, T. Jahid, I. El Affar, et al., Image analysis using separable Krawtchouk-Tchebichef's moments, in: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), May 2017, pp. 1–5.
[38] H. Karmouni, T. Jahid, A. Hmimid, M. Sayyouri, H. Qjidaa, Fast computation of inverse Meixner moments transform using Clenshaw's formula, Multimed. Tools Appl. 78 (22) (Nov. 2019) 31245–31265, https://doi.org/10.1007/s11042-019-07961-y.
[39] H. Karmouni, T. Jahid, Z. Lakhili, et al., Image reconstruction by Krawtchouk moments via digital filter, in: 2017 Intelligent Systems and Computer Vision (ISCV), Apr. 2017, pp. 1–7.
[40] M. Sayyouri, A. Hmimid, H. Qjidaa, A fast computation of Hahn moments for binary and gray-scale images, in: 2012 IEEE International Conference on Complex Systems (ICCS), Nov. 2012, pp. 1–6.
[41] T. Jahid, H. Karmouni, A. Hmimid, M. Sayyouri, H. Qjidaa, Image moments and reconstruction by Krawtchouk via Clenshaw's recurrence formula, in: 2017 International Conference on Electrical and Information Technologies (ICEIT), Nov. 2017, pp. 1–7.
[42] Y.-Q. Zhang, X.-Y. Wang, Spatiotemporal chaos in mixed linear–nonlinear coupled logistic map lattice, Phys. A, Stat. Mech. Appl. 402 (May 2014) 104–118, https://doi.org/10.1016/j.physa.2014.01.051.
[43] G. Strang, T. Nguyen, Wavelets and Filter Banks, SIAM, 1996.
[44] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (Jul. 1989) 674–693, https://doi.org/10.1109/34.192463.
[45] A.F. Nikiforov, V.B. Uvarov, S.K. Suslov, Classical orthogonal polynomials of a discrete variable, in: Classical Orthogonal Polynomials of a Discrete Variable, Springer, 1991, pp. 18–54.
[46] H. Zhu, Image representation using separable two-dimensional continuous and discrete orthogonal moments, Pattern Recognit. 45 (4) (Apr. 2012) 1540–1558, https://doi.org/10.1016/j.patcog.2011.10.002.
[47] R.M. May, Simple mathematical models with very complicated dynamics, Nature 261 (5560) (1976) 459–467.
[48] SQAM - Sound Quality Assessment Material, https://sound.media.mit.edu/resources/mpeg4/audio/sqam/. (Accessed 23 September 2020).
[49] S. Katzenbeisser, F.A.P. Petitcolas, Digital Watermarking, vol. 2, Artech House, London, 2000.
[50] P. Kabal, An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality, TSP Lab Technical Report, Department of Electrical & Computer Engineering, McGill University, 2002.
[51] TSP Lab - Software, http://www-mmsp.ece.mcgill.ca/Documents/Software/. (Accessed 24 September 2020).
[52] V. Bhat, I. Sengupta, A. Das, An audio watermarking scheme using singular value decomposition and dither-modulation quantization, Multimed. Tools Appl. 52 (2–3) (2011) 369–383.
[53] V. Stojanovic, D. Prsic, Robust identification for fault detection in the presence of non-Gaussian noises: application to hydraulic servo drives, Nonlinear Dyn. 100 (3) (2020) 2299–2313.
[54] X. Dong, S. He, V. Stojanovic, Robust fault detection filter design for a class of discrete-time conic-type non-linear Markov jump systems with jump fault signals, IET Control Theory Appl. 14 (14) (2020) 1912–1919.
[55] V. Stojanovic, S. He, B. Zhang, State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises, Int. J. Robust Nonlinear Control 30 (16) (2020) 6683–6700.
[56] L. Zhou, H. Tao, W. Paszke, V. Stojanovic, H. Yang, PD-type iterative learning control for uncertain spatially interconnected systems, Mathematics 8 (9) (2020) 1528.
[57] P. Cheng, M. Chen, V. Stojanovic, S. He, Asynchronous fault detection filtering for piecewise homogenous Markov jump linear systems via a dual hidden Markov model, Mech. Syst. Signal Process. 151 (2021) 107353.
[58] M. Yamni, H. Karmouni, M. Sayyouri, H. Qjidaa, Robust zero-watermarking scheme based on novel quaternion radial fractional Charlier moments, Multimed. Tools Appl. 80 (2021) 21679–21708, https://doi.org/10.1007/s11042-021-10717-2.
[59] M. Yamni, H. Karmouni, A. Daoui, O. El ogri, M. Sayyouri, H. Qjidaa, Blind image zero-watermarking algorithm based on radial Krawtchouk moments and chaotic system, in: 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Jun. 2020, pp. 1–7.

Mohamed Yamni was born in Fez, Morocco, in 1993. He received the B.Eng. degree in electrical engineering and the M.S. degree in engineering science from the Faculty of Science, University of Sidi Mohammed Ben Abdellah, Fez, Morocco in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in the Department of Physics, University of Sidi Mohammed Ben Abdellah, Fez, Morocco. His research interests include image watermarking, signal processing, and pattern classification.

Hicham Karmouni was born in Fez, Morocco in 1992. Received the M.S. degree in engineering science and Ph.D degrees in Signals, Systems and Informatics from the Faculty of science, University of Sidi Mohammed Ben Abdellah, Fez, Morocco in 2015 and 2019, respectively. His research interest includes, information system algorithms computing image processing pattern classification, orthogonal systems, and special functions.

Mhamed Sayyouri received the M.S. Diploma in Engineering Sciences and PhD in Signals, Systems and Computer Science from the Faculty of Sciences, University of Sidi Mohammed Ben Abdellah, Fez, Morocco in 2004 and 2014, respectively. Since 2016, he has been a research professor at the National School of Applied Sciences of the Department of Industrial Engineering, Sidi Mohamed Ben Abdellah-Fez University, Fez, Morocco. His research interests include image processing, signal and pattern classification systems.

Hassan Qjidaa received the M.S. and Ph.D. degrees in electrical engineering from Nuclear Physics Institute of Lyon, in France 1984 and 1987, respectively. Since 1987, he served as a Research Scientist at Faculty of science in the University of Sidi Mohammed Ben Abdellah, Fez, Morocco. He has been a Professor with the Department of Physics. He is currently also a Director of the Information Analysis and Microsystem Teams (IAMS), and Vice Director of the electronic signal.