

Secure Audio Watermarking for Multipurpose Defensive Applications

Salma Masmoudi¹, Maha Charfeddine¹ and Chokri Ben Amar²

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

Keywords: Audio Watermarking, Tamper Detection, Authenticity, Integrity Control, Recovery.

Abstract: Audio recordings contain very sensitive content, such as historical archival material in public archives that protects and conserves our cultural heritage, digital evidence in the context of law enforcement, the online formats of sensitive digital Holy Quran, etc. Such audio content is vulnerable to doctoring and falsification of its origin with malicious intent. One tool to solve several multimedia security difficulties facing this sensitive content is to tag it with a message before the distribution process. This technique is called watermarking. Hence, this paper aims to present a scheme of tamper detection and integrity control based on multipurpose and secure audio watermarking. To treat the integrity control application, we suggested embedding in the digital audio signal the tonal components resulting from the Human Psychoacoustic Model masking study, which are extracted as features from the relevant low-frequency band of the original audio signal. In addition, a Multilayer perceptron-based denoising autoencoder was executed after learning robust representation from corrupted audio features to correct the watermarked frequencies, thereby restoring the original ones. Consequently, blind tamper detection and blind invertibility were guaranteed. The detailed results indicated that the suggested scheme achieved higher performance at the integrity control and tamper detection level, as well as at the watermarking and reversibility properties.

1 INTRODUCTION


Audio recordings contain very sensitive content such as historical material in public archives that conserve our cultural heritage, digital evidence in the context of law enforcement, the online formats of sensitive digital Holy Quran, etc. Such audio content is vulnerable to falsification of its origin. Henceforth, the reliability and provenience of such digital audio content and the sureness about its origin are very serious factors.


To address this issue, it become essential to use a mechanism protecting and verifying the authenticity and the integrity of digital sound recordings. Ordinary watermarking techniques produced an alteration in the signal to protect that causes loss of data. Accordingly, watermarking schemes controlling the integrity of the digital content becomes compulsory. In these particular schemes, the alterations can be located in the watermarked content. However, if a sig-


nal is watermarked with an ordinary scheme for integrity control and is opposed to attacks, then its significant parts or/and the hidden watermark can disappear, and it is not possible to reconstruct it.

For this reason, it is interesting to conceive sophisticated versatile watermarking techniques for copyright protection (Masmoudi et al., 2020), integrity control, blind tamper detection (Masmoudi et al., 2024) and blind-recovering recovery that are also robust enough to compensate for the drawbacks of ordinary watermarking schemes. The blindness concept signifies that the original audio signal isn't needed either in the detection process or in the tamper detection and reversibility ones. In this paper, we introduce a multipurpose defensive audio watermarking technique based on different NN (Neural Networks) architectures and exploiting HPM (Human Psychoacoustic Model) properties with LPC (Linear Prediction Coding) envelope estimation of the audio spectral, which is original reversible, robust and blind.

This paper is planned as follows: section two presents a literature review of the recent works. Section 3 presents the multipurpose audio watermarking

^a <https://orcid.org/0000-0002-4827-2158>

^b <https://orcid.org/0000-0003-2996-4113>

^c <https://orcid.org/0000-0002-0129-7577>

system. In addition, Section 4 illustrates the experiments and results. Finally, the conclusion is presented in the last section, along with perspectives for future research.

2 RELATED WORKS

Sometimes it is indispensable to verify the authenticity of input content, i.e., to decide whether the data are original, fake, or a modified version of the original one.

Authentication techniques are the solutions to these problems. They are conceived for integrity verification and source origin authentication. They are implemented using digital signature (Romney and Parry, 2006) or digital watermarking.

The digital signature is a non-repudiation encrypted message digest extracted from the digital content. It is generally stored as a separate file which can be attached to the data to attest integrity and originality. In contrast, digital watermarking techniques insert a watermark into the digital content so that the watermark is residing in protection of this content.

In this context, fragile watermarking has been employed in the few past years to counter the dilemma of content authentication and tamper localization of image (Awasthi and Nirmal,) and video (El'Arbi et al., 2011; Tarhouni et al., 2023) hosts. The main intend of such schemes is to be fragile in content manipulation attacks; i.e. good detection performance on tamper localization while robust to conventional signal processing operations (e.g. resampling, adding noise, and filtering). But, in many application fields, such as criminal execution, the court and news, recorded audio files could be unkindly falsified or removed during transmission.

As a result, these digital signals should be checked to decide whether they are authentic or changed (Li et al., 2014; Tong et al., 2013). As the Human Auditory System (HAS) is more sensible than the Human Visual System (HVS) (Ali et al., 2022), fragile audio watermarking schemes for content authentication and tamper detection are more defiant than those for image and video. Hence, research on audio watermarking in tampering detection and recovery has been suggested in recent years.

In (Hu and Lee, 2019), the authors introduced multi-purpose audio watermarking found on Lifting Wavelet Transform LWT decomposition to fulfill authentication and copyright protection. Following the 3-level Lifting Wavelet Transform (LWT) decomposition of the audio signal, the coefficients in chosen subbands are partitioned into frames for embedding. To

enlarge applicability, the robust watermark including proprietary information, synchronization code, and frame-related data was principally hidden in the approximation subband using perceptual-based rational dither modulation (RDM) and adaptive quantization index modulation (AQIM). The fragile watermark is a highly compressed version of the embedded audio. Hashing comparison and source-channel coding make it possible to recognize tampered frames and restore affected regions. Experimentation indicates that the inserted robust watermark can endure common attacks, and the fragile watermark is very operational in tamper detection and recovery. The integration of a frame synchronization mechanism makes the suggested system endure cropping and replacement attacks. The perceptual evaluation shows that the watermark is inaudible and the scheme is appropriate for content authentication applications.

Moreover, to guarantee the requirements of the International Federation of Phonographic Industry (IFPI) for robustness, imperceptibility, payload, and audio integrity, the paper (Narla et al., 2021) proposes a robust and blind digital audio watermarking (DAW) scheme. The suggested method uses quantization index modulation to include a pre-processed watermark picture into singular values of audio signal coefficients. To detect audio tampering, such as deletion, copy-move, and substitute attacks, a hash produced using the SHA-512 method is put in watermarked audio frames. Audio is split, and then each segment is turned into a matrix to achieve tamper detection. Audio is split, and then each segment is turned into a matrix to achieve tamper detection. Each matrix segment is subjected to singular value decomposition (SVD), and the median is computed. To create the secret key, these values are logically XOR with the encrypted watermark.

In addition, (Liu et al., 2024) suggests a new authentication and recovery watermarking scheme for encrypted audio. The authors present the relative energy (RE) feature and analyze the characteristics of the feature. In the inserting process, the host audio is firstly encrypted. Then the resulting encrypted audio is split into frames. The embedding of watermark bits into each frame is done by quantifying the RE feature. At the decoding stage, the receivers locate the attacked frames based on the watermark extraction and substitute the attacked frames using 0 amplitude signals, which are scattered over different segments after anti-scrambling transformation and do not influence the expressed meaning of watermarked signal. Experimental evaluation results demonstrate that the scheme improves the security of audio signals stored on third-party servers.

3 PROPOSED METHOD

In this section, we depict an enhanced secure multipurpose audio watermarking scheme based on two different Neural Network architectures in the frequency domain able to assure very high imperceptibility, robustness, security, integrity, blind detection, blind tamper localization and blind recovering (reversibility) proprieties.

In general, robust watermarks are not affected when the watermarked data is attacked. These watermarks are often used in copyright protection applications. However, fragile watermarks can be destroyed by data manipulation, and these are also called a tamper-proof watermark. The fragile watermark can detect the modifications in the signal and also recognize the place where the modifications have occurred and also the signal before the change. Therefore, these watermarks are used for content authentication, integrity control, and tamper localization. Hence, to conceive a multipurpose audio watermarking we use a fragile-content watermarking approach (Steinebach and Dittmann, 2003) combining robust-watermarking and fragile-content features.

The adopted watermarking scheme is presented in (Charfeddine et al., 2022). It inserts imperceptibly the watermark and assures a good robustness to various attacks by exploiting BPNN architecture in the embedding and extraction processes and by studying some HPM proprieties with the LPC envelope estimation of the PSD (Power Spectrum Density). This watermarking scheme is denoted DCT-NNS-MPH.

We then explain the importance of selecting relevant features from the audio signal as watermark to ensure integrity control and after that tamper detection and recovery. We present an MLP-based denoising autoencoder (Charte et al., 2018) adopted to train the chosen original and watermarked attacked features with their indexes and positions to perform then adequately their content (feature-frequencies or features-values) reversibility.

We achieve blind frame-resynchronization in the detection process in the case of particular desynchronization attacks. This re-synchronization mechanism is blind when the proposed watermarking scheme is robust and promises then reversibility.

In the case of unrobustness, the re-synchronization mechanism is semi-blind and only the indexes of the original features in the audio frame (and the original frame numbers, when the attack is very destructive) are needed by the receiver to assure then reversibility and the original feature frequencies recovering.

Effectively, neither the original audio signal nor the real feature values are transmitted to the receiver to perform detection, tamper localization and recovery. In both types of reversibility, simulating the MLP-based denoising autoencoder DAE on the extracted features from the watermarked and attacked audio signal permits to recover of particularly relevant original frequencies by removing tampered information (denoising).

Figure 1 illustrates the general proposed audio watermarking scheme: from an audio file, a feature vector (W1) is extracted. W1 encloses W1-frames, W1-indexes and W1-values (frequencies). The embedded watermark is the concatenation of W1-frames and W1-indexes only. The watermarked audio signal is then transferred via a noisy channel. Next, the watermark (W2-frames and W2-indexes) is extracted and a newly generated feature vector (W3-frames, W3-indexes and W3-values) is extracted from the watermarked and attacked file.

If W1-frames and W2-frames are equal and W1-indexes and W2-indexes too, then robustness is achieved. If W2-frames and W3-frames are equal and W2-indexes and W3-indexes also, then authenticity is reached and integrity is preserved.

However, if W2-frames and W3-frames are different or/and W2-indexes and W3-indexes are distinct, then authenticity and integrity are not achieved. Thus, based on re-synchronization mechanism depending on the robustness of our scheme, we can perform blind tamper detection and blind invertibility by correcting tampered relevant frequency parts after simulating the MLP-based denoising autoencoder on the extracted features and without using the original audio nor W1-values.

3.1 HPM Based Relevant Features Extraction

We need to yield a binary representation of the audio content that is small enough to be hidden as a watermark and significant enough to identify alterations. To maintain audio content's semantic integrity, only a part of its full spectrum is regularly required.

For our scheme, we select a low-frequency band to collect suitable and relevant features since this band defines the most significant contents of the audio signal. Specifically, we choose pertinent tonal components resulting from the fifth step of the Global Masking Threshold Ltg algorithm of the HPM (Pan, 1995) and discard noise frequencies.

In reality, these tonal components constitute the decimated maskers that do not affect the audio quality if they are inserted later in the middle frequency

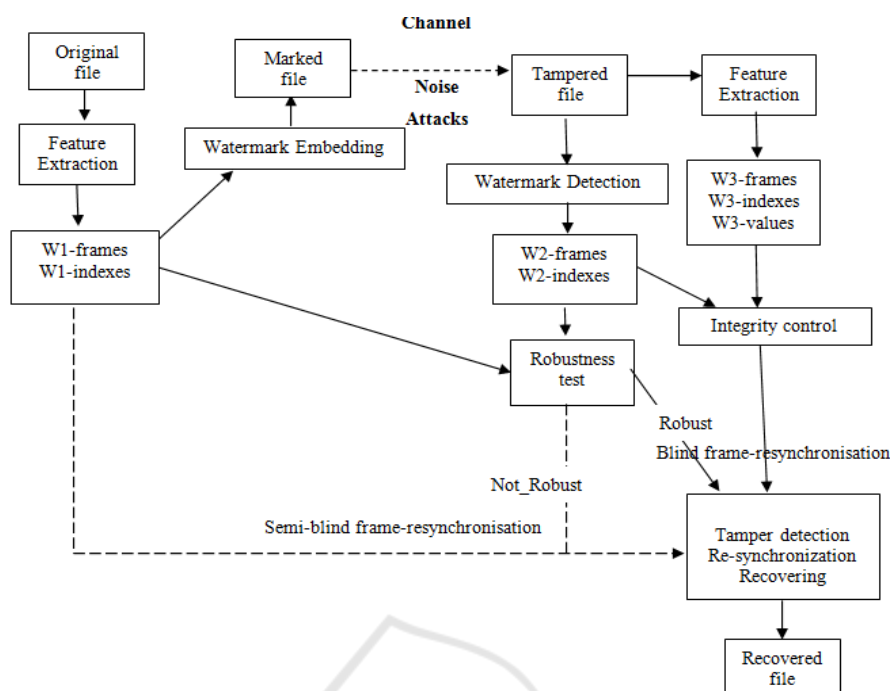


Figure 1: General scheme of the proposed tamper detection system based on DCT-NNS-MPH watermarking.

since they are situated under the Lt_g curve. In addition, decimation constitutes the adopted feature reduction approach in our proposed scheme permitting to diminish the amount of data embedded as watermark. After separating the sampled audio signal into slighter frames and dividing the signal into 32 sub-bands by a time-frequency mapping filterbank using a pseudo-Quadrature mirror filter QMF filter (Cruz-Roldan et al., 2000), the fifth step prompting to retrieve the relevant tonal features is computed:

1. FFT is calculated for the conversion from time to frequency.
2. Sound pressure level is determined in each sub-band.
3. Threshold in quiet is determined (absolute threshold of hearing).
4. Tonal or sinusoid-like components only of the audio signal are found (not considering the non-tonal ones).
5. Tonal maskers are decimated and constitute the searched pertinent features.

It is important to distinguish between tonal and non-tonal components as shown in figure 3.

For computing the global masking threshold, it is necessary to derive the tonal and non-tonal components from the FFT spectrum. This stage begins with the localization of local maxima after extracting tonal components (sinusoids-like) from low-

frequency band (4Khz) in a bandwidth of a critical band.

In our case, we don't consider the non-tonal components since they are noisy parts of the audio signal and therefore they are less significant in point of view integrity and authentication of the audio signal. In addition, ignoring them permits adequate reduction of the features.

We finally generate a vector W1 containing relevant tonal features. Each is identified by a frame number "W1-frame", an index within this frame "W1-index", and the frequency "W1-value".

3.2 Used MLP-Based Denoising Autoencoder for Tamper Detection and Reversibility

An autoencoder can be seen as the composition of an encoding map f which projects inputs onto a different feature space, and a decoding map g which operates inversely.

The main objective of the autoencoder is to recover as much information as possible from the original input, so it will attempt to minimize the distance between the inputs and the outputs. The distance function used in the cost function is usually the Mean Squared Error MSE (Charte et al., 2018), the output units should use an unbounded activation function. The quality of learned features can be evaluated by the

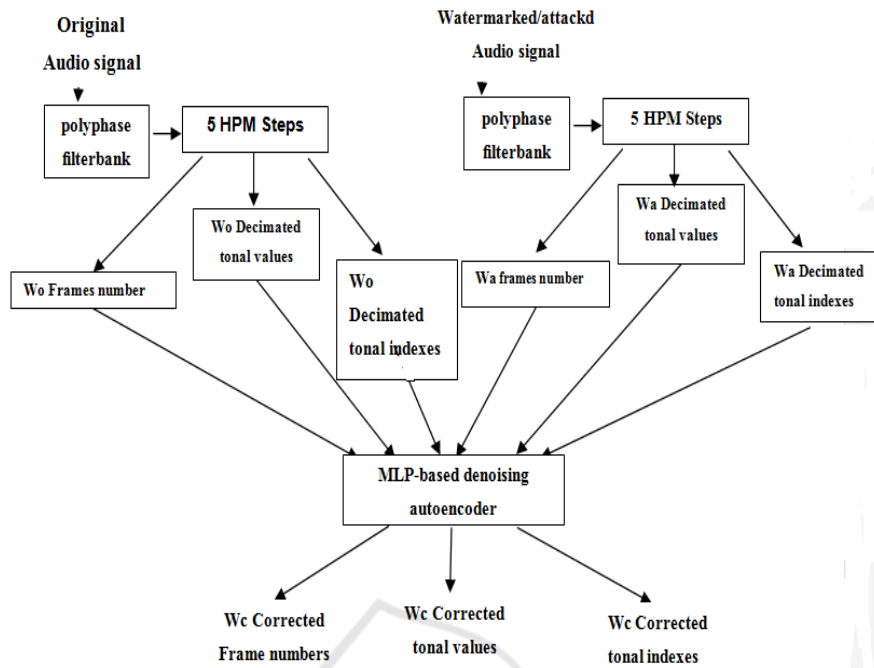


Figure 2: MLP-based denoising autoencoder training process.

model's ability to project instances back to the original feature space.

For this purpose, regression metrics can be used. Some common metrics that serve to assess the usefulness of the learned features are Mean squared error, Root Mean Squared Error, Mean absolute error and Mean absolute percentage error (Charte et al., 2018).

The proposed audio watermarking scheme uses the decimated tonal features extracted from real-life audio examples, their indexes in a selected frame, and the frame number of both the original and attacked audio signals as inputs to the DAE during the insertion process. The objective is to recover original information from tampered audio signals by removing the attacked information. The idea was inspired by the behavior of the denoising DAE (Charte et al., 2018), which attempts to learn a robust representation from corrupted data. Similarly, we consider the attacked features as input to the DAE and it is trained to recover original ones by correcting altered data as illustrated in figure 2.

In the detection process, we simulate the trained MLP-based denoising DAE with the extracted features from the watermarked attacked audio signals. We obtain as output a correction of the extracted features. Thus, we substitute the tampered content values in the low-frequency band with the corrected content values in the correspondent frames and the adequate tonal indexes (which are already re-synchronized if destructive attacks cause de-synchronization prob-

lems in addition to tampers).

Thanks to this MLP-based denoising DAE architecture with consideration of the re-synchronization process, we adequately locate the tampering and ensure the invertibility and recovery of the audio signal. We considered sparse autoencoder by activity regularization to explicitly seek an efficient learned representation (Charte et al., 2018). It is handled through an L1 penalty on the activations with a coefficient λ_1 added to the cost function during training to increase the amount of sparsity in the learned representations. Additionally, an L2 regularization or weight decay technique is used by adding a penalty term with coefficient λ_2 to the cost function.

The training is based on the backpropagation algorithm using the Adam optimizer with an initial learning rate 0.005 and 3500 epochs. During the training phase, a learning rate scheduler was considered to reduce with drop-based technique the learning rate. This prevents the gradient descent from sticking into local minima.

4 EXPERIMENTAL RESULTS

This section is dedicated to presenting several experiments carried out to test the performance characteristics of this proposed audio watermarking method.

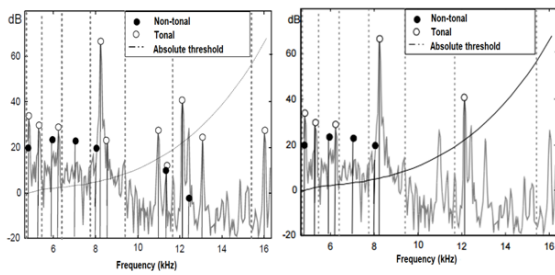


Figure 3: Discrimination between tonal and non-tonal components.

4.1 Testing Environment

In this work, different MATLAB simulations are performed. MP3 compression and audio StirMark attacks are common transformations used in the literature by scientific researchers, which can have a significant impact on the robustness of the watermark, the integrity of the audio signal, and the precision of the detection process. For the compression operation, we used standard tools such as the lame Audio Encoder. For other audio manipulations, we used the standard StirMark Benchmark for Audio (SMBA) tool with default parameters (WVL, 2006) and Audacity 2.3.3. Experimental tests are performed on original WAVE audio files of type music and Quranic-sensitive audio signals. Owing to the sensitivity of Quranic verses, there is a crucial need to incessantly monitor Quranic verses dispatched through Internet websites to make sure that they are not changed or fraudulent and are authenticated.

All audio signals have 44.1 KHz as the sampling rate, 16 bits as several bits per sample, and a duration of around 20 s. We present the result of a selection of some audio signals which each one is threatened to 49 Stirmark attacks and three MP3 compression attacks with three different bitrates. The selected WAVE audio files are musical audio and Quranic-sensitive signals.

The average length of the extracted features constituting the watermark is about 2100 bits depending on the sound duration. NC and BER are calculated to evaluate the similarity between the extracted original features and the inserted ones (or between the detected watermark and the attacked extracted features).

4.2 Inaudibility Results

Transparency performance ensures that the watermarking scheme does not degrade the host signal significantly. Otherwise, the watermark embedding process did not introduce distinguishable noise in the host carrier. The objective difference grade (ODG)

(Acedo, 2006) measure is used. ODG can take a value between -4 and 0 . The closer the value of ODG to 0 , the more degradation is imperceptible.

Table 1 shows the inaudibility results of the proposed scheme.

Due to the exploitation of the frequency perceptual masking associated with the LPC estimation of the digital audio signal, SNR values are significantly higher than the designed value by the IFPI (20 dB) (Eya et al., 2013).

In addition, we observe that all the ODG values are less than -1 . Thus, the proposed scheme satisfies the inaudibility requirements of optimal audio watermarking techniques.

4.3 Robustness Results

Figure 4 exhibits the MP3 and Stirmark robustness results. From this figure and the results presented in the paper (Charfeddine et al., 2022), we observe very good MP3 robustness results (even, with 64Kbps as compression rate which is a very destructive attack). Thus, we realize that using the HPM in the frequency domain assures not only perfect inaudibility but also good robustness to MP3 compression.

When observing figure 4 and based on the results in the paper (Charfeddine et al., 2022), we deduce that the DCT-NNS-HPM scheme has good robustness results except for some destructive attacks with highly damaging perceptive effects.

Experimental results have revealed that the exploitation of frequency perceptual masking studied in HPM with the spectral envelope concern in the frequency domain is very interesting with very good inaudibility and robustness results.

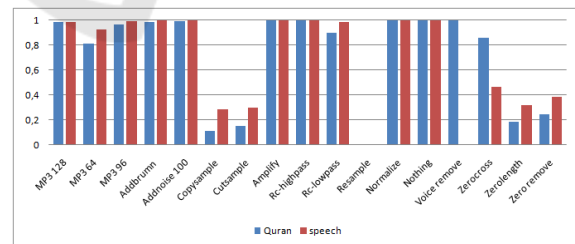


Figure 4: Robustness results of the DCT-NNS-HPM watermarking scheme.

4.4 Comparison

Inaudibility and robustness comparison results with the previous scheme (Maha et al., 2010) and other published audio watermarking schemes are presented in (Charfeddine et al., 2022) and in tables 2 and 3.

Table 1: Inaudibility results of the DCT-NNS-HPM scheme.

File \ Metrics	Speech1	Speech2	Speech3	Quran 1	Quran 2	Quran 3
SNR	38,0132	45,8644	39,368	42,1998	44,3242	42,2308
ODG	-0,9249	-0,9512	-0,3992	-0,7589	-0,4725	-0,6626

Table 2: Average ODG comparison with other watermarking methods.

Inaudibility	(Xiang et al., 2015)	(Xue et al., 2019)	(Korany et al., 2023)	(Hu and Lee, 2019)	The proposed Method
Absolute Average ODG	1.1623	0.6175	0.245	[0.56, 0.79]	0.43

4.5 Integrity Control Results

To check whether the watermarked audio signal is tampered or not after Stirmark attacks, we verify integrity by comparing the original content features and the extracted content of the attacked watermarked audio file.

If no attack occurs, the bit rate error (BER) is equal to zero. When observing table 4, we show test results after performing MP3 attacks with different bit rates and Stirmark attack also.

For example in table 4, the attacks MP3 with 128Kbps, invert, extra stereo and adbrumn superior to 5100 present results BER equal or inferior to the nooperation attack “nothing”. An error rate equal or below the bit error of the nooperation attack can be seen as a threshold for verifying integrity. Content signal processing like the removal of voice and removal of samples has higher error rates than the nooperation attack as they are very destructive.

4.6 Tamper Detection Results

In the tamper localization experimental results, we begin by verifying the integrity of the audio signal. As we explained previously, tamper detection can be preceded also by a re-synchronizing mechanism to adjust the frame positions or/and corresponding indexes of the extracted features from tampered audio signals. Attacks can cause de-synchronization problems in addition to tampers.

Consequently, we can deduce when observing table 5, that if integrity is OK (BER equal or inferior to the nooperation attack “nothing”), neither re-synchronization nor tamper localization, DAE denoising and recovering are needed. However, if integrity fails (NO), then it is important to check the robustness propriety to decide on the re-synchronization mechanism type. In effect, if the robustness is achieved but the integrity fails, so blind re-synchronization is executed.

Thus, tamper localization constitutes the frame positions or/and their corresponding indexes of the detected watermark (due to the robustness propriety)

and is followed by the DAE simulation and the recovery of the real value features. However, when robustness failed and integrity also, then semi-blind re-synchronization is compulsory depending on only the received frame positions or/and corresponding indexes (without needing the original real value features) which constitute the tampered parts of the audio signal that will be after that recovered thanks to DAE denoising.

4.7 Reversibility Results

The same tamper localization conditions are applied to the recovery process and depend on the robustness and then the re-synchronization process type. Thus after simulating the DAE on the tampered features, this system proceeds to perform denoising and correcting the altered data. Getting the denoised features as the output of the DAE, we compare them with the original features by calculating the BER:

- If BER is equal or inferior to the nooperation attack “nothing”, then perfect recovering is achieved, we obtain consequently similar audio file to the watermarked signal without attacks.
- If BER equal or inferior to 0.2, then satisfactory recovering is accomplished, we obtain partially comparable audio signal to original one.

Finally if BER superior to 0.2 then recovering is failed and it is not possible to attain even a partially similar audio signal to the original file.

5 CONCLUSION

In this paper, we have introduced a reversible, robust and blind NNS-based audio watermarking scheme exploiting HPM masking proprieties of MPEG audio standard and benefiting from the advantages of LPC envelope estimation of the audio spectral density. Imperceptibility and robustness are accomplished thanks to the exploitation of a BPNN architecture in the insertion and detection processes with consideration of HPM masking benefits and also LPC envelope estimation advantages.

Table 3: Average BER comparison with other watermarking methods.

Attacks	(Xiang et al., 2015)	(Xue et al., 2019)	(Korany et al., 2023)	(Hu and Lee, 2019)	The proposed Method
Resampling	0,53	0,5913	0,4676	0	0
AddNoise	3,4835	5,0429	2	0,61	0
Amplify	0,0184	0,0398	0,0199	—	0
MP3 (128 kbps)	0,0184	0,0797	0,017	0,06	0,013
HighPassFilter	0,0184	0,0398	0,0376	—	0
LowPassFilter	0,0184	0,0398	0,039	0	0,01

Table 4: Integrity results of the proposed tamper detection based on DCT-NNS-HPM watermarking for a sensitive signal.

Attack Integrity	Rob Dec	Ber.frame	Ber.index	Ber.reel	Int Dec
MP3 128	Robust	0,007	0,040	0,087	NO
MP3 64	Robust	0,040	0,040	0,120	NO
MP3 96	Robust	0,020	0,013	0,093	NO
Addbrunn 1	Robust	0,013	0,053	0,120	NO
Addbrunn 2	Robust	0,100	0,100	0,233	NO
Addbrunn 3	Robust	0,533	0,427	0,253	NO
Addbrunn 4	Robust	0,013	0,000	0,167	NO
Addbrunn 5	Robust	0,020	0,007	0,220	NO
Addbrunn 6	Robust	0,033	0,007	0,233	NO
Addbrunn 7	Robust	0,033	0,007	0,233	NO
Addbrunn 8	Robust	0,053	0,020	0,233	NO
Addbrunn 9	Robust	0,053	0,033	0,233	NO
Addbrunn 10	Robust	0,073	0,047	0,233	NO
Addbrunn 11	Robust	0,100	0,053	0,233	NO
Addffnoise	Non Robust	1,000	0,913	0,000	NO
Addnoise 1	Robust	0,013	0,060	0,100	NO
Addnoise 2	Robust	0,033	0,080	0,133	NO
Addnoise 3	Robust	0,047	0,060	0,127	NO
Addnoise 4	Robust	0,100	0,040	0,153	NO
Addnoise 5	Robust	0,107	0,047	0,160	NO
Addsinus	Robust	0,007	0,073	0,120	NO
Amplify	Robust	0,000	0,000	0,107	NO
Compressor	Robust	0,173	0,033	0,167	NO
Copysample	Non Robust	0,713	0,120	0,213	NO
Cutsample	Non Robust	0,647	0,187	0,200	NO
Dynnoise	Robust	0,147	0,027	0,147	NO
Echo	Non Robust	0,613	0,067	0,253	NO
Exchange	Robust	0,007	0,053	0,100	NO
Extrastereo 30	Robust	0,000	0,000	0,033	OK
Extrastereo 50	Robust	0,000	0,000	0,033	OK
Extrastereo 70	Robust	0,000	0,000	0,033	OK
FFT hipass	Robust	0,027	0,060	0,127	NO
FFt Invert	Non Roust	0,000	0,000	0,027	NO
FFt Real reverse	Robust	0,007	0,073	0,120	NO
FFt Stat1	Non Robust	0,380	0,087	0,187	NO
FFt test	Non Robust	0,380	0,087	0,200	NO
Flipsample	Non Robust	0,287	0,067	0,167	NO
Invert	Non Robust	0,000	0,000	0,033	OK
Lsbzero	Robust	0,007	0,073	0,107	NO
Normalize	Robust	0,007	0,073	0,087	NO
Nothing	Robust	0,000	0,000	0,033	OK
Original	Robust	0,000	0,000	0,033	OK
Rc-highpass	Robust	0,227	0,040	0,207	NO
Rc-lowpass	Robust	0,007	0,053	0,107	NO
Resample	Robust	1,000	0,000	0,767	NO
Smooth	Robust	0,027	0,047	0,080	NO
Smooth 2	Robust	0,060	0,067	0,153	NO
Stat1	Robust	0,000	0,000	0,040	NO
Stat2	Robust	0,007	0,053	0,100	NO
Voice remove	Non Robust	1,000	0,000	0,767	NO
Zerocross	Robust	0,040	0,080	0,147	NO
Zerolength	Non Robust	0,480	0,060	0,173	NO
Zero remove	Non Robust	0,240	0,067	0,173	NO

In addition, authentication and integrity are adequately controlled thanks to the appropriate choice of the extracted features to be hidden in the audio signal. These features constituting the relevant tonal coefficients are extracted from the significant low-frequency band of the cover audio signal after HPM study.

Concerning blind detection, blind tamper detection, and blind-recovery processes, they are accom-

Table 5: Tamper detection results of the proposed tamper detection based on DCT-NNS-HPM watermarking for sensitive signals.

tamper detection Attack	Rob Dec	Int Dec	Sync	DAE	Tamp detect
MP3 128	Robust	NO	Blind	OK	OK
MP3 64	Robust	NO	Blind	OK	OK
MP3 96	Robust	NO	Blind	OK	OK
Addbrunn 100	Robust	NO	Blind	OK	OK
Addbrunn10100	Robust	NO	Blind	OK	OK
Addbrunn 1100	Robust	NO	Blind	OK	OK
Addbrunn 2100	Robust	NO	Blind	OK	OK
Addbrunn 3100	Robust	NO	Blind	OK	OK
Addbrunn 4100	Robust	NO	Blind	OK	OK
Addbrunn 5100	Robust	NO	Blind	OK	OK
Addbrunn 6100	Robust	NO	Blind	OK	OK
Addbrunn 7100	Robust	NO	Blind	OK	OK
Addbrunn 8100	Robust	NO	Blind	OK	OK
Addbrunn 9100	Robust	NO	Blind	OK	OK
Addffnoise	Non Robust	NO	Semi-blind	OK	OK
Addnoise 100	Robust	NO	Blind	OK	OK
Addnoise 300	Robust	NO	Blind	OK	OK
Addnoise 500	Robust	NO	Blind	OK	OK
Addnoise 700	Robust	NO	Blind	OK	OK
Addnoise 900	Robust	NO	Blind	OK	OK
Addsinus	Robust	NO	Blind	OK	OK
Amplify	Robust	NO	Blind	OK	OK
Compressor	Robust	NO	Blind	OK	OK
Copysample	Non Robust	NO	Semi-blind	OK	OK
Cutsample	Non Robust	NO	Semi-blind	OK	OK
Dynnoise	Robust	NO	Blind	OK	OK
Echo	Non Robust	NO	Semi-blind	OK	OK
Exchange	Robust	NO	Blind	OK	OK
Extrastereo 30	Robust	OK	-	-	-
Extrastereo 50	Robust	OK	-	-	-
Extrastereo 70	Robust	OK	-	-	-
FFT hipass	Robust	NO	Blind	OK	OK
FFt Invert	Non Roust	NO	Semi-blind	OK	OK
FFt Real reverse	Robust	NO	Blind	OK	OK
FFt Stat1	Non Robust	NO	Semi-blind	OK	OK
FFt test	Non Robust	NO	Semi-blind	OK	OK
Flipsample	Non Robust	NO	Semi-blind	OK	OK
Invert	Non Robust	OK	Semi-blind	OK	OK
Lsbzero	Robust	NO	Blind	OK	OK
Normalize	Robust	NO	Blind	OK	OK
Nothing	Robust	OK	-	-	-
Original	Robust	OK	-	-	-
Rc-highpass	Robust	NO	Blind	OK	OK
Rc-lowpass	Robust	NO	Blind	OK	OK
Resample	Robust	NO	Blind	OK	OK
Smooth	Robust	NO	Blind	OK	OK
Smooth 2	Robust	NO	Blind	OK	OK
Stat1	Robust	NO	Blind	OK	OK
Stat2	Robust	NO	Blind	OK	OK
Voice remove	Non Robust	NO	Semi-blind	OK	OK
Zerocross	Robust	NO	Blind	OK	OK
Zerolength	Non Robust	NO	Semi-blind	OK	OK
Zero remove	Non Robust	NO	Semi-blind	OK	OK

plished by using an MLP-based denoising autoencoder associated with a re-synchronization mechanism. The DAE based-re-synchronization mechanism permits during simulation in the detection stage to delete noises and correct tampers even after desynchronization problems, attacks or audio signal manipulations. Copyright protection, authentication, integrity control, blind tamper detection and blind invertibility are reached efficaciously. A comparable

MLP-based denoising DAE training architecture can be performed on the original frequencies and those enclosing the watermark (with possible alterations due to some attacks) which permits to correcting the watermarked frequencies after DAE simulation in the detection process and reconstructing the original ones. The goal here is to reconstruct completely the host signal, either its significant parts from which we have extracted the features and also its recovered original coefficients that are modified after watermarking. This total recovery can be very interesting in some applications where the cover signal must be entirely recuperated.

REFERENCES

- Acevedo, A. G. (2006). Audio watermarking quality evaluation. In *E-business and Telecommunication Networks*, pages 272–283. Springer.
- Ali, B. Q. A., Shahadi, H. I., Kod, M. S., and Farhan, H. R. (2022). Covert voip communication based on audio steganography. *International Journal of Computing and Digital Systems*, 11(1):821–830.
- Awasthi, A. and Nirmal, M. L. Multiple image watermarking approach based on anfis for copyright protection and image authentication: A review.
- Charfeddine, M., Mezghani, E., Masmoudi, S., Amar, C. B., and Alhumyani, H. (2022). Audio watermarking for security and non-security applications. *IEEE Access*, 10:12654–12677.
- Charte, D., Charte, F., García, S., del Jesus, M. J., and Herrera, F. (2018). A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Information Fusion*, 44:78–96.
- Cruz-Roldan, F., Lopez-Ferreras, F., Amo-Lopez, P., and Maldonado-Bascon, S. (2000). Pseudo-qmf filter banks for audio signals subband coding. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 1, pages 237–240. IEEE.
- El'Arbi, M., Charfeddine, M., Masmoudi, S., Koubaa, M., and Amar, C. B. (2011). Video watermarking algorithm with bch error correcting codes hidden in audio channel. In *IEEE symposium series in computational intelligence*, pages 164–17.
- Eya, M., Maha, C., and Chokri, B. A. (2013). Audio silence deletion before and after mpeg video compression. In *2013 International Conference on Computer Applications Technology (ICCAT)*, pages 1–5. IEEE.
- Hu, H.-T. and Lee, T.-T. (2019). Hybrid blind audio watermarking for proprietary protection, tamper proofing, and self-recovery. *IEEE Access*, 7:180395–180408.
- Korany, N. O., Elboghdady, N. M., and Elabdein, M. Z. (2023). High capacity, secure audio watermarking technique integrating spread spectrum and linear predictive coding. *Multimedia Tools and Applications*, pages 1–24.
- Li, J., Wang, R., Yan, D., and Li, Y. (2014). A multipurpose audio aggregation watermarking based on multistage vector quantization. *Multimedia tools and applications*, 68:571–593.
- Liu, Z., Cao, Y., and Lin, K. (2024). A watermarking-based authentication and recovery scheme for encrypted audio. *Multimedia Tools and Applications*, 83(4):10969–10987.
- Maha, C., Maher, E., Mohamed, K., and Chokri, B. A. (2010). Dct based blind audio watermarking scheme. In *2010 International conference on signal processing and multimedia applications (SIGMAP)*, pages 139–144. IEEE.
- Masmoudi, S., Charfeddine, M., and Ben Amar, C. (2020). A semi-fragile digital audio watermarking scheme for mp3-encoded signals using huffman data. *Circuits, Systems, and Signal Processing*, 39:3019–3034.
- Masmoudi, S., Charfeddine, M., and Ben Amar, C. (2024). Mp3 audio watermarking using calibrated side information features for tamper detection and localization. *Multimedia Tools and Applications*, pages 1–26.
- Narla, V. L., Gulivindala, S., Chanamallu, S. R., and Gangwar, D. (2021). Bch encoded robust and blind audio watermarking with tamper detection using hash. *Multimedia Tools and Applications*, 80(21-23):32925–32945.
- Pan, D. (1995). A tutorial on mpeg/audio compression. *IEEE multimedia*, 2(2):60–74.
- Romney, G. W. and Parry, D. W. (2006). A digital signature signing engine to protect the integrity of digital assets. In *2006 7th International Conference on Information Technology Based Higher Education and Training*, pages 800–805. IEEE.
- Steinebach, M. and Dittmann, J. (2003). Watermarking-based digital audio data authentication. *EURASIP Journal on Advances in Signal Processing*, 2003:1–15.
- Tarhouni, N., Masmoudi, S., Charfeddine, M., and Amar, C. B. (2023). Fake covid-19 videos detector based on frames and audio watermarking. *Multimedia Systems*, 29(1):361–375.
- Tong, X., Liu, Y., Zhang, M., and Chen, Y. (2013). A novel chaos-based fragile watermarking for image tampering detection and self-recovery. *Signal Processing: Image Communication*, 28(3):301–308.
- WVL, D. (2006). Audio benchmarking tools and steganalysis. Technical report, Technical report, ECRYPT-European Network of Excellence in Cryptology.
- Xiang, Y., Natgunanathan, I., Rong, Y., and Guo, S. (2015). Spread spectrum-based high embedding capacity watermarking method for audio signals. *IEEE/ACM transactions on audio, speech, and language processing*, 23(12):2228–2237.
- Xue, Y., Mu, K., Li, Y., Wen, J., Zhong, P., and Niu, S. (2019). Improved high capacity spread spectrum-based audio watermarking by hadamard matrices. In *Digital Forensics and Watermarking: 17th International Workshop, IWDW 2018, Jeju Island, Korea, October 22-24, 2018, Proceedings 17*, pages 124–136. Springer.