HIGH HIDING CAPACITY AUDIO WATERMARKING METHOD BASED ON DISCRETE COSINE TRANSFORM

Omar Abodena[1], Ali Alashtir[b]
[1]Department of Computer Science, Faculty of Science Gharyan, University of Gharyan, Libya
[2]Department of Computer, Tripoli College of Science and Technology, Libya

ABSTRACT

This paper presents a new high-capacity algorithm for audio watermarking for the purpose of embedding a watermark audio signal into an original audio signal. This algorithm uses DCT (discrete cosine transform) in combination with SVD (singular value decomposition), DWT (discrete wavelet transform), and CZT (chirp z-transform). For the purpose of ensuring high levels of imperceptibility and robustness, a low-frequency sub band of 1-level DWT is selected to embed the watermark, followed by DCT, CZT, and SVD. As a result, the proposed algorithm achieves a high capacity of 352800 bits per second while a good quality audio signal is maintained (having an objective difference grade of more than -0.04). It also provides high imperceptibility (a signal to noise ratio of more than 58 dB) and it is extremely resistant to common audio attacks, including requantizing, resampling, additive white Gaussian noise or AWGN, MP3 compression, and low-pass filtering. The proposed algorithm presents much better compared to the audio watermarking algorithms already existent.

Keyword:- *Chirp z-transform, Singular value decomposition, Discrete wavelet transform, Discrete cosine transform, Audio watermarking*

1. INTRODUCTION

Digital watermarking represents a highly effective approach to authentication and copyright protection. This process embeds copyright information into original media, whether in the form of audio, video or image [1]. Digital audio watermarking has four primary requirements for efficacy [2]: security, data capacity, robustness and imperceptibility. Algorithms for watermarking can be generally divided into three groups. These classifications are: blind algorithms, where the process of watermark extraction does not require the host audio signal; semi-blind algorithms, which require some information about the host audio signal; and non-blind algorithms, which require the host audio signal in its entirety [3]. A review of existing literature reveals a number of proposed algorithms for digital watermarking of audio, all of which aim to achieve high levels of imperceptibility and robustness, as well as a large data capacity. These algorithms include DWT (discrete wavelet transform) [4], [5], QR decomposition [6], [7], DCT (discrete cosine transform) [8], [9], SVD (singular value decomposition) [10], [11], and lastly, spread spectrum [12], [13]. Al-Haj A. (2014) [14] proposed a non-blind algorithm for audio watermarking which utilized both SVD and DWT in cascade form, applying 3-level DWT to segmented audio signals first, and following this up with an application of SVD.

Mohsenfar S.M. et al. (2015) [6] proposed a highly robust algorithm for watermarking which used QR decomposition and genetic algorithm. In this algorithm, the original audio was divided into several frames, after which each segment was exposed to QR and decomposed; the genetic algorithm then identified the optimal place within every frame. A novel scheme for blind audio watermarking was devised by Hu H.T. et al. (2015) [9] this scheme achieved a high embedding capacity and excellent robustness by modulating vectors within the DCT domain to an auditory masking constraint in order to perform the process of embedding the watermark. A further algorithm for watermarking audio was proposed by Xiang Y. et al. (2015) [13] this method utilized spread spectrum, meaning that it was able to obtain higher capacity while also ensuring good imperceptibility and robustness.

This paper presents and develops a novel algorithm for non-blind hybrid watermarking of audio, based on chirp z-transform (CZT), singular value decomposition (SVD), discrete wavelet transform (DWT) and discrete cosine transform (DCT). The algorithm proposed possesses a far higher embedding capacity than other methods, while ensuring good robustness and imperceptibility.

The primary contributions of this algorithm, as developed in this paper, are as follows:
(i) The audio signal is used as a watermark to be embedded.
(ii) A stereophonic sound is used in both the original and the watermark audio signal in order to achieve high embedding capacity.
(iii) This algorithm achieves a much higher embedding capacity of 352800 bits per second.
(iv) To ensure satisfactory imperceptibility and robustness, the benefits of the hybrid transform domain, including DCT, CZT, DWT, and SVD, are utilised.

Section 2 of this paper represents a background review of the subject. Section 3 describes the proposed algorithm for audio watermarking. Section 4 presents the experimental results achieved and discusses these, while section 5 provides the subsequent conclusions.

2. BACKGROUND REVIEW

1.1 Discrete Wavelet Transform

The primary advantages of DWT are its multi-resolution analysis and good localization properties; it represents the closest simulation of the human auditory system (HAS), according to theoretical models [15]. In DWT, audio signals are decomposed into four separate sub bands: high-high (HH); high-low (HL), low-high (LH) and low frequency (LL). Because of its multi resolution analysis, it is possible to embed a watermark into any of these bands [16]. In general, however, watermarks embedded into the high frequency band will have higher levels of imperceptibility, but may have compromised robustness. Meanwhile, watermarks embdded into low frequency sub bands will more effectively be able to resist various attacks [15], [16].

1.2 Discrete Cosine Transform

DCT, or discrete cosine transform, is a method of increasing robustness in watermarking algorithms. In DCT, an input signal is decomposed into three separate sub bands: low frequency, middle frequency, and high frequency. Characteristic of DCT is its ‘energy compaction’ property: in this method, the majority of the signal’s significant information can be concentrated into only a few of the DCT’s coefficients. The important parts of a signal are contained in the high-energy, low-frequency sub-bands, whereas the high frequency sub bands are more resistant to audio attacks [9], [17].

1.3 Chirp Z-transform

The chirp z-transform algorithm, or CZT, can be used in order to compute the z-transforms of sequences of samples. With this algorithm, the z-transform can be evaluated efficiently at various points along the z-plane, which lies on spiral or circular contours, commencing at any arbitrary point on the plane [18], [19]. It is possible to represent z-domain transfer functions as polynomials, for which zeros and poles serve as the roots. The behavior of a system can be effectively analysed by describing the system according to its zeros and poles. The poles here represent the roots of the transfer function’s feedback element, and the zeroes the roots of its feed forward element. The primary advantage of utilizing chirp z-transform over other algorithms is that it is possible to enhance the sharpness of resonance peaks by computing a z-transform across a contour lying closer to the pole or poles. Chirp z-transform can also greatly improve the frequency resolution of the frequency spectrum, once analysed, by zooming it [20], [21].

1.4 Singular Value Decomposition

Singular value decomposition is a mathematical tool for linear algebra, used to analyse matrices, which is extensively utilised in the watermarking technique [22-24]. Through the use of SVD, matrix $\boldsymbol{A}$ of size m × n can be decomposed into three separate matrices as follows:

$$A = USV^T$$

In this equation, U is a m × m orthogonal matrix, S is a m × n diagonal matrix with positive elements and V is a n × n orthogonal matrix. Matrix $\boldsymbol{A}$*’s* singular values, or SVs, make up the elements of *S*. SVs have some attractive properties [25], [26], such as:

i) Changing SVs does not affect signal quality.
ii) The SVs are invariant under various attacks.
iii) The SVs satisfy effective algebraic advantages.

3. THE PROPOSED ALGORITHM FOR AUDIO WATERMARKING

This section describes the process of watermark extracting and watermark embedding. So as to ensure high robustness and imperceptibility of the algorithm under higher embedding capacity, the low-frequency sub-band of 1-level DWT is selected as the means of embedding the watermark, to be followed by DCT, CZT, and SVD.

3.1 Embedding Watermark Procedure

The procedure for watermark embedding is illustrated in block diagram form in Fig. 1, below. The detailed steps for the embedding watermark procedure are:

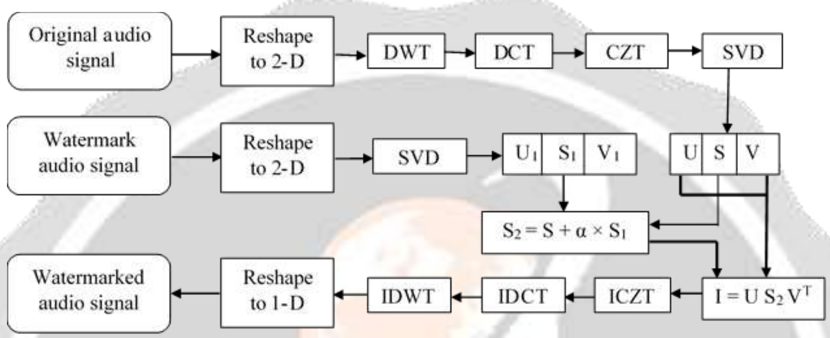

Fig -1: Block diagram of the watermark embedding procedure

Step 1. Reshape the 1-D original audio signal into a 2-D audio signal.

Step 2. Apply 1-level DWT to the 2-D original audio signal A to decompose it into four separate sub bands, as follows:

$$LL\ LH\ HL\ HH = DWT(A)$$

Step 3. Calculate DCT of low-frequency sub-band LL as follows:

$$D = DCT(LL)$$

Step 4. Compute CZT of previous step as follows:

$$C = CZT(D)$$

Step 5. Apply SVD to C in order to decompose it further, as follows:

$$[U\ S\ V] = SVD(C)$$

Step 6. Reshape the 1-D watermark audio signal into the 2-D audio signal.

Step 7. Apply SVD to the 2-D watermark audio signal W as follows:

$$[U_1 S_1 V_1] = SVD(W)$$

Step 8. Estimate the new singular value by adding the decomposed singular value of the original audio signal to the singular value of the watermarked audio signal, multiplied by a scaling factor of α. This factor α determines the added watermarked audio signal's strength, as follows:

$$S_2 = S + (\alpha \times S_1)$$

Step 9. Combine orthogonal matrices from the original audio signal in its decomposed form with the new singular value:

$$I = U\ S_2\ V^T$$

Step 10. Compute inverse CZT of I as follows:

$$CZ = ICZT(I)$$

Step 11. Calculate inverse DCT of CZ so that a watermarked sub band of low frequency is obtained, as follows:

$$LL_1 = IDCT(CZ)$$

Step 12. Apply inverse DWT in order to obtain the audio signal in watermarked form. Instead of LL, modified LL_1 is utilised as follows:

$$AA = IDWT\ (LL_1 LH\ HL\ HH)$$

Step 13. Reshape the 2-D watermarked audio signal into the 1-D watermarked audio signal.

3.2 Extracting Watermark Procedure

The procedure of extracting a watermark is illustrated in block diagram form in Fig. 2, below. The detailed steps for the procedure are:

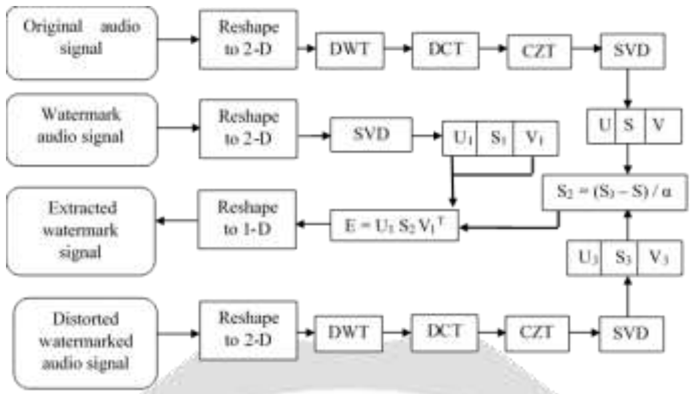

Fig -2: Block diagram of the extracting watermark procedure

Step 1. Reshape the 1-D original audio signal into the 2-D audio signal.

Step 2. Apply 1-level DWT to the 2-D original audio signal, *A*, to decompose it into four separate sub bands as follows:

$$LL\ LH\ HL\ HH = DWT(A)$$

Step 3. Calculate DCT of low-frequency sub-band *LL* as follows:

$$D = DCT(LL)$$

Step 4. Compute CZT of previous step as follows:

$$C = CZT(D)$$

Step 5. Apply SVD to *C* in order to decompose it further, follows:

$$[U\ S\ V] = SVD(C)$$

Step 6. Reshape the 1-D watermarked audio signal into the 2-D audio signal.

Step 7. Apply 1-level DWT to the 2-D watermarked audio signal W_A to decompose it into four sub-bands as follows:

$$LL_1\ LH_1\ HL_1\ HH_1 = DWT(W_A)$$

Step 8. Calculate DCT of low-frequency sub-band LL_1 as follows:

$$D_1 = DCT(LL_1)$$

Step 9. Compute CZT of previous step as follows:

$$C_1 = CZT(D_1)$$

Step 10. Apply SVD to C_1 in order to decompose it further, as follows:

$$[U_3\ S_3\ V_3] = SVD(C_1)$$

Step 11. Reshape the 1-D watermark audio signal into the 2-D audio signal.

Step 12. Apply SVD to the 2-D watermark audio signal *W* as follows:

$$[U_1 S_1 V_1] = SVD(W)$$

Step 13. Estimate the new singular value by subtracting decomposed singular value of the original audio signal from that of the watermarked audio, dividing by scaling factor of α in order to obtain the extracted watermark audio signal's singular value, as follows:

$$S_2 = (S_3 - S)\ /\ \alpha$$

Step 14. Combine orthogonal matrices from the watermark audio signal with the extracted watermark audio signal S_2*'s* singular value to obtain the extracted watermark audio signal, as follows:

$$E = U_1\ S_2\ V_1^T$$

Step 15. Reshape the 2-D extracted watermark audio signal into the 1-D extracted watermark audio signal.

4. EXPERIMENTAL RESULTS AND SUBSEQUENT DISCUSSION

The experiments have been developed in a MATLAB R2012b environment on different music styles, including rock, jazz, pop, and speech as original audio signals. Each music is 16 bits/sample stereo audio signal in the WAVE format sampled at 44.100 kHz with a duration of 51.2 s. The embedded music watermark "piano" is a 16-bit stereo audio signal sampled at 44.100 kHz, in WAVE format, with a duration of 12.8 s.

4.1 Imperceptibility Test

The objective difference grade (ODG) and signal to noise ratio (SNR) are two approaches that are widely used to perform the perceptual quality assessment.
SNR [27] can be utilised to measure a difference indicator between the watermarked and the original audio signals, which is calculated as follows:

$$SNR(dB) = 10log_{10} \frac{\sum_{i=1}^{l} x_i^2}{\sum_{i=1}^{l} \left(x_i - x_i'\right)^2}$$

where x_i' and x_i are the watermarked and the original audio signal, respectively.

Objective difference grade (ODG) [28] is utilised in order to establish the quality of a watermarked audio signal. The grading standards of ODG are presented in Table (1), below.
The software EAQUAL [29], which evaluates the quality of audio, is used to obtain the ODG values. The imperceptibility results obtained for the proposed algorithm are demonstrated in Table (2). As can be seen in this table, the proposed algorithm has SNR values of higher than 58 dB. This means that the algorithm performs better on the issue of imperceptibility.

Table -1: ODG objective grading standard

Grade	Description	Quality
0.00	Imperceptible	Excellent
-1.00	Perceptible, but not annoying	Good
-2.00	Slightly annoying	Fair
-3.00	Annoying	Poor
-4.00	Very annoying	Bad

Table -2: SNR and ODG between original and watermarked audio signal

Audio file	SNR	ODG
Rock	64.465	0.03
Jazz	59.143	0.03
Pop	63.123	0.03
Speech	58.262	-0.04

The proposed algorithm has ODG values of greater than -0.04, which means that the proposed algorithm has better perceptual quality.

4.2 Robustness Test

To establish the effectiveness and robustness of the algorithm proposed, a number of audio attacks were performed upon the audio signals after watermarking, utilising a MATLAB R2012b environment and software Adobe Audition 3.0. These attacks consisted of:

- **Additive white Gaussian noise (AWGN):** The watermarked signal receives added Gaussian white noise with an SNR of 55dB.
- **Requantizing:** The audio signal, once watermarked, is first requantized from the original 16-bit/sample to an 8-bit/sample, and afterwards is requantized again to a 16-bit/sample.
- **Resampling:** The watermarked audio signal is resampled to 11.025 and 8.000 kHz, after which it is sampled back to 44.1 kHz.
- **Low-pass filtering:** The watermarked audio signal receives low-pass filtering with a cutoff frequency of 22.05 kHz.
- **MP3 compression:** MPEG-1 layer 3 compression is applied to the audio signal after watermarking in order to transform WAVE format into MP3, after which it is transformed back again. The compression rates 64 and 32 kbps/channel are utilised.

Normalised cross-correlation [30] is utilised in order to determine levels of similarity between the extracted and the original watermark audio signal. It is defined as follows:

$$NC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} W(i,j)\acute{W}(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} W^2(i,j)}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} \acute{W}^2(i,j)}}$$

Table (3) summarises the performance of the proposed algorithm in terms of robustness when subjected to the common attacks detailed above. As the table indicates, the proposed algorithm has higher NC values for various music styles against different attacks. All NC values are 1 for all music styles, except AWGN attack for rock and speech music styles, which means the algorithm is highly resistant to these attacks. From these results, it is clear that the algorithm is extremely robust.

Table -3: NC of extracted watermark from different music styles

Attack type	Rock	Jazz	Pop	Speech
No Attack	1.000	1.000	1.000	1.000
Resampling 11.025	1.000	1.000	1.000	1.000
Resampling 8.000	1.000	1.000	1.000	1.000
Requantizing	1.000	1.000	1.000	1.000
AWGN	0.999	1.000	1.000	0.905
MP3 32 Kbps	1.000	1.000	1.000	1.000
MP3 64 Kbps	1.000	1.000	1.000	1.000
Low-pass filtering	1.000	1.000	1.000	1.000

4.3 Capacity

Capacity refers to the number of watermark bits that may reliably be embedded into an original audio signal within a certain period of time, where time is expressed as a unit [31]. It is computed [22] as follows:

$$Capacity \ = \frac{N_w}{Time}(bps)$$

$Time$ here represents the original audio signal's duration, and N_w indicates the number of bits it is possible to embed into the original audio signal. For computing the audio signal's bits [32], the number of bits is calculated as follows:

$$File\ size\ (in\ bits) = \ sample\ rate\ \times bit\ depth\ \times number\ of\ channels\ \times length\ (in\ seconds)$$

The algorithm's capacity for data embedding is 352800 bps for all applied music styles.

Table (4) gives a comparison of the algorithm proposed with other contemporary algorithms for watermarking, in terms of how each performed on capacity, SNR, and MP3. As is clearly demonstrated in the table, the proposed algorithm has better SNR and excellent robustness against MP3 compression, while it also maintains a higher embedding capacity in comparison to other audio watermarking algorithms. The proposed algorithm meets the required criteria of robustness and imperceptibility for watermarking procedures.

Table -4: Comparison of the proposed algorithm's performance against that of other algorithms for audio watermarking

Algorithm	Capacity	SNR (dB)	MP3
Bhat, V. et al. (2010) [26]	45.9	22.11-26.84	1.000 (32 kbps)
Khaldi K. and Boudraa A. (2013) [33]	50.3	24.12.26.38	1.000 (32 kbps)
Wang, X. et al. (2013) [15]	102.4	26.41	1.000 (32 kbps)
Mohsenfar S.M. et al. (2015) [6]	159	24.62-27.33	0.943 (128 kbps)
Al-Haj A. (2014) [14]	1387	40.14-51.26	0.976 (32 kbps)
Proposed	352800	58.26-64.46	1.000 (32 kbps)

5. CONCLUSION

This paper has proposed a novel algorithm for high embedding watermarking of audio, which, by comparison to competing algorithms, makes better use of SVD, DWT, DCT and CZT in order to achieve its goals. In the interest of high data embedding capacity, a stereo audio signal used as a watermark is embedded into 1-level DWT where a low-frequency sub-band is selected to obtain good robustness and high imperceptibility. Experimental simulation results have indicated that the proposed audio watermarking algorithm permits a higher embedding capacity than other algorithms, while also maintaining the audio signal's quality at a high level. Simultaneously, the algorithm is extremely robust against common audio attacks, including resampling, MP3 compression, requantizing, low-pass filtering, and AWGN.

6. REFERENCES

[1] Liang, X., Xiang, S., (2020), Robust reversible audio watermarking based on high-order difference statistics. Signal Processing, **173**, 107584.

[2] Hua, G., Huang, J., Shi, Y.Q., Goh, J., Thing, V.L., (2016), Twenty years of digital audio watermarking—a comprehensive review., Signal Processing, **128**, 222-242.

[3] Akhaee, M.A., Kalantari, N.K., Marvasti, F., (2010), Robust audio and speech watermarking using Gaussian and Laplacian modeling. Signal processing, **90**(8), 2487-2497.

[4] Huang, H.N., Chen, S.T., Lin, M.S., Kung, W.M., Hsu, C.Y., (2015), Optimization-based embedding for wavelet-domain audio watermarking., Journal of Signal Processing Systems, **80**(2), 197-208.

[5] Abodena, O., Agoyi, M., (2018), Colour image blind watermarking scheme based on fast walsh hadamard transform and hessenberg decomposition., Studies in Informatics and Control, **27**(3), 339-348.

[6] Mohsenfar, S.M., Mosleh, M., Barati, A., (2015), Audio watermarking method using QR decomposition and genetic algorithm., Multimedia Tools and Applications, **74**(3), 759-779.

[7] Kumar, A., Singh, A., Prakash, S., Singh, V., (2021), A novel approach towards audio watermarking using FFT and CORDIC-Based QR Decomposition., AI and IoT-Based Intelligent Automation in Robotics, 323-338.

[8] Hu, H.T., Hsu, L.Y., Chou, H.H., (2014), Perceptual-based DWPT-DCT framework for selective blind audio watermarking., Signal Processing, **105**, 316-327.

[9] Hu, H.T., Hsu, L.Y., (2015), Robust, transparent and high-capacity audio watermarking in DCT domain., Signal Processing, **109**, 226-235.

[10] Abdelwahab, K.M., Abd El-atty, S.M., El-Shafai, W., El-Rabaie, S., Abd El-Samie, F.E., (2020), Efficient SVD-based audio watermarking technique in FRT domain., Multimedia Tools and Applications, **79**(9), 5617-5648.

[11] Rezaei, A., Khalili, M., (2019), A robust blind audio watermarking scheme based on DCT-DWT-SVD., In: Fundamental Research in Electrical Engineering, 101-113.

[12] Li, R., Xu, S., Yang, H., (2016), Spread spectrum audio watermarking based on perceptual characteristic aware extraction., IET Signal Processing, **10**(3), 266-273.

[13] Xiang, Y., Natgunanathan, I., Rong, Y., Guo, S., (2015), Spread spectrum-based high embedding capacity watermarking method for audio signals., IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), **23**(12), 2228-2237.

[14] Al-Haj, A., (2014), A dual transform audio watermarking algorithm., Multimedia tools and applications, **73**(3), 1897-1912.

[15] Wang, X., Wang, P., Zhang, P., Xu, S., Yang, H., (2013), A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform., Signal Processing, **93**(4), 913-922.

[16] Rasti, P., Samiei, S., Agoyi, M., Escalera, S., Anbarjafari, G., (2016), Robust non-blind color video watermarking using QR decomposition and entropy analysis., Journal of Visual Communication and Image Representation, **38**, 838-847.

[17] Khalili, M., (2015), DCT-Arnold chaotic based watermarking using JPEG-YCbCr., Optik-International Journal for Light and Electron Optics, **126**(23), 4367-4371.

[18] Bagchi, S., Mitra, S.K., (1999), The Nonuniform Discrete Fourier Transform and Its Applications in Signal Processing.

[19] Gomez, C., Bunks, C., Chancelier, J.-P., Delebecque, F., Goursat, M., Nikoukhah, R., Steer, S., (1999), Engineering and Scientific Computing with Scilab.

[20] Rabiner, L., Schafer, R., Rader, C., (1969), The chirp z-transform algorithm., IEEE transactions on audio and electroacoustics, **17**(2), 86-92.

[21] Rabiner, L.R., Schafer, R.W., Rader, C.M., (1969), The Chirp z-Transform Algorithm and Its Application., Bell System Technical Journal, **48**(5), 1249-1292.

[22] Lei, B., Soon, Y., Zhou, F., Li, Z., Lei, H., (2012), A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition., Signal Processing, **92**(9), 1985-2001.

[23] Abodena, O., Agoyi, M., Celebi, E., (2017), Hybrid technique for robust image watermarking using discrete time fourier transform., In: 25th Signal Processing and Communications Applications Conference (SIU), 1-4.

[24] Arora, S.M., (2018), A DWT-SVD based robust digital watermarking for digital images., Procedia computer science, **132**, 1441-1448.

[25] Bhat, V., Sengupta, I., Das, A., (2011), An audio watermarking scheme using singular value decomposition and dither-modulation quantization., Multimedia Tools and Applications, **52**(2-3), 369-383.

[26] Bhat, V., Sengupta, I., Das, A., (2010), An adaptive audio watermarking based on the singular value decomposition in the wavelet domain., Digital Signal Processing, **20**(6), 1547-1558.

[27] Chen, C.-J., Huang, H.-N., Tu, S.-Y., Lin, C.-H., Chen, S.-T., (2021), Digital audio watermarking using minimum-amplitude scaling on optimized DWT low-frequency coefficients., Multimedia Tools and Applications, **80**(2), 2413-2439.

[28] Khalil, M., Adib, A., (2014), Audio watermarking with high embedding capacity based on multiple access techniques., Digital Signal Processing, **34**, 116-125.

[29] Lerch, A., (2002), Zplane development, EAQUAL-Evaluate Audio QUALity.

[30] Su, Q., Chen, B., (2018), Robust color image watermarking technique in the spatial domain., Soft Computing, **22**(1), 91-106.

[31] Fallahpour, M., Megías, D., (2015), Audio watermarking based on Fibonacci numbers., IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), **23**(8), 1273-1282.

[32] Waller, D., Weidmann, A., GCSE Computer Science for OCR Student Book. Cambridge University Press, 2016.

[33] Khaldi, K., Boudraa, A., (2013), Audio watermarking via EMD., IEEE transactions on audio, speech, and language processing, **21**(3), 675-680.