Adaptive Segmentation on Audio Watermarking using Signal Differential Concept and Multibit Spread Spectrum Technique

1st Rozan Nauf Firmansyah
School of Electrical Engineering
Telkom University
Bandung, Indonesia
rozanauf@gmail.com

2nd Gelar Budiman
School of Electrical Engineering
Telkom University
Bandung, Indonesia
gelarbudiman@telkomuniversity.ac.id

3rd Sofia Sa'idah
School of Electrical Engineering
Telkom University
Bandung, Indonesia
sofiasaidahsfi@telkomuniversity.ac.id

2022 8th International Conference on Science and Technology (ICST) | 978-1-6654-5622-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICST56971.2022.10136289

***Abstract*—Audio data digital piracy is a crucial matter since it jeopardizes the original contents copyright. Spread spectrum (SS) audio watermarking is an effective method of preventing digital audio piracy. Numerous studies have been published about these spread spectrum (SS) audio watermarking technologies. However, those technologies did not have enough capacity and high robustness at the same time. Therefore, the audio watermarking system in this paper is based on a unique Adaptive Segmentation and the multibit SS-based that can embed multiple watermark bits into the host audio signal utilizing one random pseudo-noise (PN) to represent multiple watermark bits. By determining the highest response value of the second-order derivative of the original audio signal, the audio watermark in this paper is robust against typical assaults and has good performance against desynchronization attacks. The audio watermarking technique in this paper has high imperceptibility, has a large embedding capacity, good robustness, and has been proven to be effective through numerous simulations.**

***Keywords*—Audio watermarking, adaptive segmentation, digital audio piracy, multi-bit spread spectrum.**

I. Introduction

Downloading and uploading multimedia files has increased dramatically in recent years as a result of the internets rapid expansion. A massive amount of multimedia information is easily accessible to individuals. Furthermore, digital creations in the form of images, sounds, and videos, among other digital formats, have been published on the internet [1]. Digital works are becoming more vulnerable to piracy over time since pirates will always find a method to exploit the easiness with which digital works are being made public, distributed, processed, and duplicated for the purpose of violating the lawful rights of copyright holders for selfish enrichment [2]. Due to prevent digital piracy, digital publishers must secure their intellectual property when creating multimedia material. The unauthorized use, distribution, and illegal selling are referred to as digital piracy [3]. Digital watermarking is one potential method for preventing digital piracy.

Digital watermarking is the technique of hiding data by embedding it into host signals such as video, audio, and photos, though this study focused solely on audio data. Digital audio watermarking can then be used to establish rights of ownership, ensure authorized access, prevent unlawful reapplications, and simplify content authentication [4]. In audio watermarking, information for copyright protection can be subtly inserted into an audio file. For example, copyright information for music files such as publisher ID, user ID, and file transaction details, or even for the medical industry file where the hospital can insert their watermark into cardiac signals. An authorized individual or organization may use a secret key to obtain the watermark data from a watermarked audio file without utilizing the original audio file. Three parameters are used to assess the efficacy of audio watermarking [5]. These include imperceptibility, robustness, and embedding capacity. The majority of digital audio publishers want to include digital watermarks in their works for copyright protection and integrity certification without affecting audio quality [6]. Therefore, audio watermarking must be effective and useful.

Spread Spectrum (SS) [7] – [10], echo hiding [11] – [12], patchwork [13] – [14], and others [15] – [16]. have all been proposed as audio watermarking approaches. Researchers have concentrated on SS-based audio watermarks in these methods because the embedding and extraction structures are simple and capable of performing exceptionally well in terms of reliability against common signal processing intrusions while maintaining imperceptibility and embedding capacity [8].

The principle of SS-based audio watermarking is the watermark bits of information are dispersed over the host audio signal spectrum using reference patterns generated by pseudo-random sequence producers. Traditional SS-based audio algorithms have weak robustness against desynchronization strikes which can significantly decrease the extraction of watermark precision.

In this work, the concept of signal differential is applied to extract feature points. The feature detected points are also the main focus of the extracted and embedded segments. The Stationary Wavelet Transform (SWT) approach is used to transform the host audio signal from the time domain to the frequency domain, and the watermark is then

978-1-6654-5622-7/22/$31.00 ©2022 IEEE

embedded using the Multibit SS algorithm, depending on the decomposition level and sub-band selected.

The fundamentals of the watermarking system will be explained in the next Section II. In the Section III describes the watermark's system model. The Section IV shows the experimental results. The research's conclusions are provided in Section V.

II. Basic Theory

A. Adaptive Segmentation

The ability to achieve adaptive segmentation through using signal differential concept determines the highest response value of the second-order derivative of the original audio signal. Before obtaining the feature points, the second-order derivative of the host audio signal must be computed. The original audio signal represents by $f(x)$, the first-order derivative function symbolized by $f'(x)$, and its second-order derivation function presented by $f''(x)$). Robust Feature Points (RFPS) method in [10] is proposes a similar concept to strengthen the watermark robustness against desynchronization attacks. The following is the signal differential formula:

$$f'(x) = \frac{\partial i}{\partial x} = f(x+1) - f(x), \tag{1}$$

$$f''(x) = \frac{\partial i'}{\partial x'} = f'(x+1) - f'(x), \tag{2}$$

where the original audio sample points number is indicated by the symbol N, and the value of $x = 1, 2, ..., N-1$. The audio clip response identified by the audio feature point detector in 3 is represented by the second-order derivative function.

$$S(x) = (f''(x))^2. \tag{3}$$

The highest response points were retrieved as feature points from equation (3). The following criteria are used to locate all of the feature points in the audio clip:

$$(O - L)/2 > 1) \cap (O + L/2 - 1 < N). \tag{4}$$

where L defines the duration of the audio segment between the detected feature points and O defines the extracted feature points. This criterion must be achieved in order to eliminate feature points from the beginning and end of the original signal, allowing for more space for the watermark to be embedded.

B. Stationary Wavelet Transform

Watermarking using the frequency domain is supposed to be more robust and imperceptible. Other frequency domain methods include the Discrete Cosine Transform (DCT) [17] and Discrete Wavelet Transform (DWT) [18], both of which generate effective watermarking results. However, these techniques might function badly when the feature locations are relocated as a result of cropping and pitch invariant Time Scale Modification (TSM) results. The SWT algorithm has become one of the transformation algorithms developed to compensate for DWTs lack of translation invariance. SWT approach [19] was employed in this work to address the problem:

$$x_i(n) = \sum_{k=-\infty}^{\infty} x(k) h_i(n-k), \tag{5}$$

where $x_i(n)$ is the result of SWT output in the i-th subband, $x(n)$ is the audio signal, and $h(n)$ is the i-th subband wavelet coefficient. The signal can be analyzed using a technique known as decomposition. By transferring the information signal to the HPF and LPF, decomposition is used to get signals containing high and low frequencies. The audio signal can be rebuilt using the wavelet coefficient, which is also known as the Inverse SWT (I-SWT).

$$x(n) = \frac{1}{2}\left(\sum_{k=-\infty}^{\infty} x_1(k) h_1(n-k) + \sum_{k=-\infty}^{\infty} x_2(k) h_2(n-k) \right). \tag{6}$$

where $x_1(n)$ and $x_2(n)$ are the signals resulting from SWT decomposition in the first and second subbands, respectively. The wavelet coefficients in the first and second subbands are the $h_1(n)$ and $h_2(n)$, respectively. If a multilevel SWT decomposition is performed, the number of subbands in SWT can be calculated through the following formula:

$$N_s = N_d + 1. \tag{7}$$

where N_s is the total amount of subbands on SWT multilevel and N_d is the decomposition level in SWT. The following is the SWT output signal formulation for multilevel decomposition:

$$x^{(i)}(n) = x(n) * h^{(i)}(n). \tag{8}$$

where i is the subband index, $h^{(i)}(n)$ is the total filter in the i subband and $x^{(i)}(n)$ is SWT output on the i subband.

C. Multibit SS Audio Watermarking

The distinction between traditional SS-based audio watermarking and Multibit SS audio watermarking is that in SS-based audio watermarking, on PN sequence represents one watermark bit. In Multibit SS, a single PN sequence represents multiple watermark bits. In a brief, the multibit SS method embeds an N-bit watermark represented by p within the $L_c = 2^N$ into the host data x_0, generating the watermarked signal. The embedding method on multibit SS is suggested by [19], when $\mathbf{p}_{sj}^{(i)}$ is declared to be a random code at index j and at segments i , x_i and x_{wi} the insertion equation is as follows:

$$x_{wi} = \mathbf{x}_i + \alpha \mathbf{p}_{sj}^{(i)}. \tag{9}$$

where, in the case of multiple bits, random code $\mathbf{p}_{sj}^{(i)}$ is the outcome of mapping with the following function:

$$\mathbf{p}_{sj}^{(i)} = f_m(\mathbf{w}_j^{(i)}). \tag{10}$$

where $f_m()$ is a function that maps each possible index of multiple bits to a random code and $\mathbf{w}_j^{(i)} = \{w_1^{(i)}, w_2^{(i)}, ..., w_{N_s}^{(i)}\}$ is the multiple bits at the j index of N bits in the i-th segment. The random code correlation is not only done in one code for the extraction process of multibit SS, yet this involves the complete code because the code index must be discovered first to offer the maximum correlation result so that the digital watermark can be determined based on the maximum correlation result. All potential random codes are listed in $\mathbf{p_s} \in \mathbb{R}^{L_k \times N_p}$ below:

$$\boldsymbol{p_s} = \begin{bmatrix} \mathbf{p}_{s1} \\ \mathbf{p}_{s2} \\ . \\ . \\ . \\ \mathbf{p}_{sLk} \end{bmatrix}. \tag{11}$$

then the following formula can be used to get the index with the greatest correlation.

$$\hat{j} = \underset{j \in \{1,2,...,L_k\}}{argmax} \left| \mathbf{x}_{wi} \mathbf{p}_\mathbf{s}^\mathbf{T} \right|. \tag{12}$$

where $\hat{j}$ is the index where the result of the correlation between the indicated audio and the $\mathbf{p_s}$ code reaches its highest value. The following digital watermark will be extracted using the index j in the i-th segment, namely,

$$\hat{\mathbf{w}}_{\hat{j}}^{(i)} = \left\{ \hat{w}_{\hat{1}}^{(i)}, \hat{w}_{\hat{2}}^{(i)}, ..., \hat{w}_{\hat{N}_s}^{(i)} \right\}. \tag{13}$$

III. System Model and Design

The main operations of this system are the watermark embedding and extraction procedures. The host audio signal is in *.wav format, while the watermark is in *.bmp image format.

A. Watermark Embedding

The block diagram of watermark embedding process is depicted in Figure 1.

Step 1: Generate a random PN, then select a representative PN. The image watermark is encoded as a binary sequence.
Step 2: Calculate the 2nd order derivative of the host to obtain the feature points.
Step 3: Extract the feature points and select audio segments.
Step 4: SWT method is applied, then determines the subband.
Step 5: Multibit SS technique is used to embed the image watermark information into the selected subband.
Step 6: Inverse SWT is deployed to recover the watermarked signal.

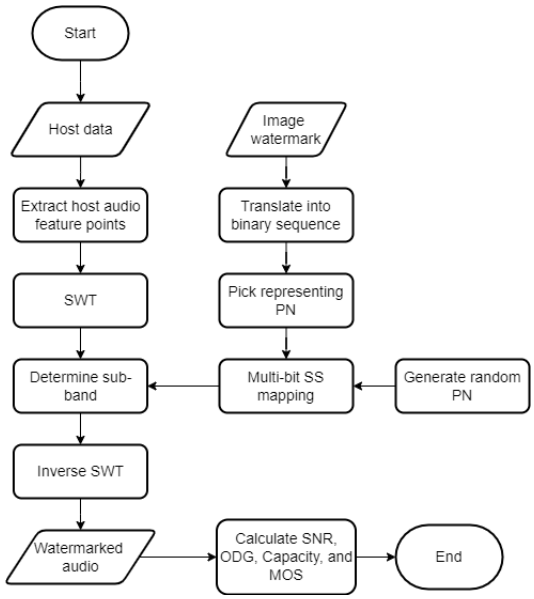

Figure 1. The procedure flow diagram for inserting a watermark

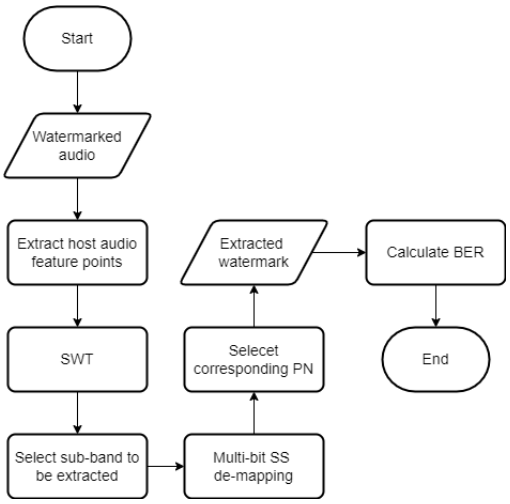

Figure 2. Block diagram for the process of extracting a watermark

B. Watermark Extraction

The watermark extraction process block diagram is depicted in Figure 2.

Step 1: The feature points are obtained by deriving the host audio signal twice.
Step 2: The feature points are used to perform adaptive segmentation.
Step 3: Choose the SWT transform subband to be extracted.
Step 4: De-mapping with a multibit SS method, followed by the selection of the corresponding PN.
Step 5: The extracted watermark is obtained.

IV. Simulation Results

In this section, the watermark is a 41×41 picture in *.bmp format, and the capacity of embedded classified info can be expanded by modifying the value to adjust the

audibility of the embedded watermark. In the following experimental results, the suggested method is applied to three different types of audio clips in WAV format, 44.1 kHz sampling frequency, mono channel, 16 bits/sample, 10 seconds duration, and at the same 172 bps embedding rate. The simulations are performed on the MATLAB and Windows 11 operating systems.

A. *Imperceptibility Performance Evaluation*

The watermarked audio is tested to both an objective and subjective quality test to assess its imperceptibility. The objective quality test is derived from Objective Different Grade (ODG) [20] and Signal-to-Noise Ratio (SNR). If the SNR value is more than 20 dB, the audio watermarking achieves its high quality. When the SNR value is low, the embedded watermark in the audio can be easily identified, otherwise, the embedded watermark is difficult to detect. The following is SNR formula.

$$SNR = 10log_{10}\frac{\sum_n x^2(i)}{\sum_n [y(i) - x(i)]^2}, \tag{14}$$

where $x(i)$ represents the host audio signal and $y(i)$ represents the watermarked audio signal The subjective quality is calculated using the Objective Difference Grade (ODG), as indicated in Table I.

TABLE I
THE SCALES OF MOS & ODG

Score	Audio quality	Description
0	Excellent	Imperceptible
-1	Good	Perceptible but annoying
-2	Fair	Slightly annoying
-3	Poor	Annoying
-4	Bad	Very annoying

The perceptual evaluation of audio quality (PEAQ) technique is used to evaluate the perceptual quality of the proposed watermarking approach [21]. The PEAQ algorithm returns an objective difference grade (ODG) parameter that ranges from -4 to 0 after comparing the host audio signal's quality with that of its watermarked variant. As the ODG value rises, the perceptual quality also does.

TABLE II
THE PERCEPTUAL QUALITY EVALUATION

Host Audio Clip	Jazz	Folk	Bass	Pop
SNR (dB)	43.28	41	50.18	44.06
ODG	-0.75	-0.71	-0.65	-0.73
Average SDG	4.46	4.5	4.5	4.6

Table II's audio chip is played to thirty listeners, who are then asked to rate each watermarked audio signal, using Table I as reference. The trade-off between imperceptibility and robustness must be handled throughout the simulation process. Table II displays the SNR and ODG averages for all audio samples. According to the experimental data, the MOS values are 4.5 on average, the ODG values are -0.7, which is suitable for watermarking, and the SNR values are 44.63 dB on average. The simulation results show that the difference between the original and watermarked audio was minimal. The watermark embedding has no effect on the audio quality.

TABLE III
THE COMPARISON OF ODG, SNR, AND CAPACITY

Parameters	**Method in [11]**	**Method in [12]**	**Proposed method**
ODG	-0.7	-0.9	-0.7
SNR	**N/A**	26.82 dB	44.63 dB
Capacity	84 bps	11.2 bps	172 bps

The perceptual quality and capacity of the proposed method is compared to the methods in [9], [10]. Based on Table III, the proposed method achieved performance improvements, indicated by the value of ODG, SNR, and capacity. This shows that the proposed method has better performance in imperceptibility and embedding capacity. The process of audio watermark embedding and extracting is illustrated in figures below:

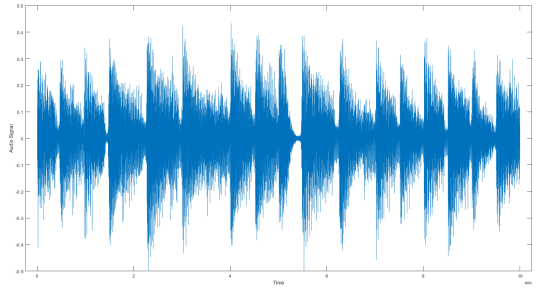

Figure 3. Audio sample before watermark embedding

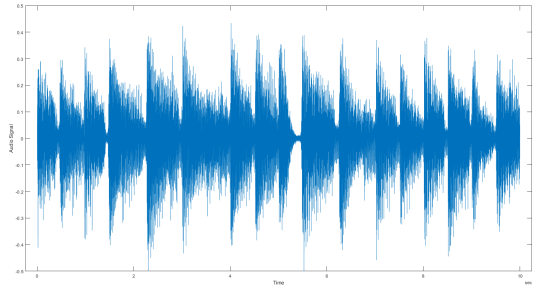

Figure 4. Audio sample after watermark embedding

The two figures, Figure 3 and Figure 4, show that the audio samples are indistinguishable between before and after watermark extraction because the audio watermark ODG is set to at least -0.7. In another case, if the ODG is set higher it can be seen the difference, see in Figure 5.

The signal wave in Figure 5 has higher audio amplitude than the audio in Figure 4. Figure 5 is the audio sample after watermark extraction performed with lower ODG. The ODG value is set to -1.17 in the simulation, producing the audio sample in Figure 5 above.

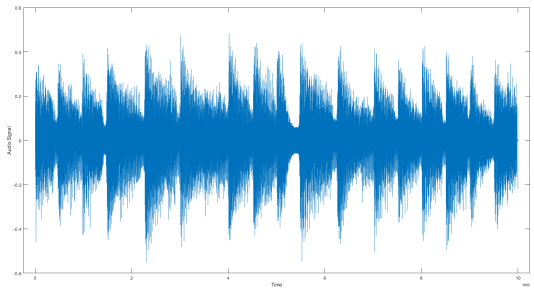

Figure 5. Audio sample after watermark embedding with lower ODG

B. Robustness Performance Evaluation

In order to evaluate the proposed system's robustness, the performance of the recommended audio watermarking technique under common signal processing attacks is tested using the Bit Error Rate (BER).

$$BER = \left(\frac{Total\ errors}{Total\ number\ of\ bits}\right) \times 100 \qquad (15)$$

The following common audio signal attacks are used for robustness evaluation:

- *Closed-loop attack* : The watermark is derived from the host audio without any deployed attacks.
- *Re-quantization attack* : The watermarked signals are re-quantized from 16 bits to 8 bits for each sample.
- *Gaussian white noise attack (AWGN)* : The watermarked signals are treated with extra white Gaussian noise respectively, with watermarked signal to noise ratios of 30 dB and 10 dB.
- *MP3 attack* : MPEG 1 Layer III compression is performed on the watermarked signals, with compression bit rates of 96 and 64 kbps, respectively.
- *Low-pass filtering (LPF)* : A low-pass filter with a cut-off frequency of 9 kHz is applied to the watermarked signals.
- *Band-pass filtering (BPF)* : A high-pass filter with a cut-off frequency of 100 Hz is deployed to the watermarked signals.

TABLE IV
THE BER (%) UNDER COMMON ATTACKS.

Attacks	**BER(%)**		
	Method in [9]	Method in [10]	Proposed method
Closed-loop	0	0	0
Re-quantization	0	0	0
Noise (10 dB)	0	0	0
Noise (30 dB)	0	0	0
MP3 (64 kbps)	0	0	0
MP3 (96 kbps)	0	0	0
High-pass filtering (100 Hz)	0	0	0
Low-pass filtering (9 kHz)	0	0	0

Table IV displays the BER of techniques when exposed to common attacks. According to the simulation results, the methods in [9], [10], and the proposed technique can accurately extract the embedded watermark bits.

In addition to typical signal assaults, the following desynchronization attacks are employed for robustness performance evaluation:

- 10% CF attack: Delete the sample points of 10% from the original audio clip
- 10% CM attack: Delete 10% of the original audio sample points in the center.
- 5% TSM: The duration of the watermarked audio signal is expanded by 5% while the original signals pitch remains unchanged.
- 15% TSM: The watermarked audio signal's duration is expanded by 15% while the original signals pitch remains unchanged.

TABLE V
THE BER (%) UNDER DESYNCHRONIZATION ATTACKS.

Attacks	**BER (%)**		
	Method in [9]	Method in [10]	Proposed method
10% CF	42.6	0	0
10% CM	50	0	8.47
5% pitch invariant TSM	44.64	6.25	16.90
15% pitch invariant TSM	58.04	15.18	27.99

Table V demonstrates that the suggested technique is more robust against desynchronization attacks than the method in [9] and has lower robustness than the method in [10], but is still within a reasonable range.

V. CONCLUSION

Due to the internets rapid expansion, copyright infringement has become a critical concern. The audio watermarking system described in this paper offers a large watermark capacity, great imperceptibility, and good robustness to typical audio signal processing attacks. The feature points were retrieved using a signal differential adaptive segmentation method, and the audio segment was then centered on the feature point. The host audio signal was converted to the frequency domain using the redundant and shift-invariant SWT technique, and the multibit SS algorithm was used for the embedding process.

Based on the simulation results, the Objective Difference Grade (ODG) is -0.7, the Signal-to-Noise Ratio (SNR) is 44.63 dB, the capacity is 172 bps, the average Bit Error Rate (BER) is 0% for common attacks, and it is 23% under desynchronization attacks. The simulation results showed that the proposed method has a substantial embedding capacity, is robust to desynchronization attacks, and can withstand conventional audio signal processing attacks. In the future, a blind watermarking technique can be applied to improve its robustness, its imperceptibility and develop a relatively similar embedding watermark approach with a huge embedding capacity as well.

ACKNOWLEDGEMENT

This research is supported by the Indonesia Minister of Education and Culture funding no. 126/SP2H/RT-MONO/LL4/2022;359/PNLT3/PPM/2022.

REFERENCES

[1] R. Subhashini and K. B. Bagan, "Robust audio watermarking for monitoring and information embedding," 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), 2017, pp. 1-4.

[2] Z. Meng, T. Morizumi, S. Miyata and H. Kinoshita, "Design Scheme of Copyright Management System Based on Digital Watermarking and Blockchain," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 2018, pp. 359-364.

[3] N. Husin and A. Hidayanto, "Impact of Piracy on Music Sales in Digital Music Transformation - A Systematic Literature Review," 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), 2018, pp. 592-596.

[4] W. Weina, "Digital audio blind watermarking algorithm based on audio characteristic and scrambling encryption," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2017, pp. 1195-1199.

[5] Chincholkar, Yugendra, S.R. Ganorkar and Kude, Shalaka, "A Survey: Digital Audio Watermark Designed Methods," 2017 IJARCCE, 6, 288-292.

[6] Xingyuan Liang and Shijun Xiang, "Robust reversible audio watermarking based on high-order difference statistics", Signal Processing, Volume 173, 2020, 107584, ISSN 0165-1684.

[7] H. S. Malvar and D. A. F. Florencio, "Improved spread spectrum: a new modulation technique for robust watermarking," in IEEE Transactions on Signal Processing, vol. 51, no. 4, pp. 898-905, April 2003.

[8] J. Mayer, "Improved Spread Spectrum multibit watermarking," 2011 IEEE International Workshop on Information Forensics and Security, 2011, pp. 1-6.

[9] Y. Xiang, I. Natgunanathan, D. Peng, G. Hua and B. Liu, "Spread Spectrum Audio Watermarking Using Multiple Orthogonal PN Sequences and Variable Embedding Strengths and Polarities," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 3, pp. 529-539, March 2018.

[10] W. Lu, L. Li, Y. He, J. Wei and N. N. Xiong, "RFPS: A Robust Feature Points Detection of Audio Watermarking for Against Desynchronization Attacks in Cyber Security," in IEEE Access, vol. 8, pp. 63643-63653, 2020.

[11] S. Wang, W. Yuan, J. Wang and M. Unoki, "Inaudible Speech Watermarking Based on Self-compensated Echo-hiding and Sparse Subspace Clustering," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2632-2636.

[12] S. Wang, W. Yuan and M. Unoki, "Multi-Subspace Echo Hiding Based on Time-Frequency Similarities of Audio Signals," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2349-2363, 2020.

[13] Z. Liu, Y. Huang and J. Huang, "Patchwork-Based Audio Watermarking Robust Against De-Synchronization and Recapturing Attacks," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 5, pp. 1171-1180, May 2019.

[14] Y. Chincholkar and S. Ganorkar, "Audio Watermarking Algorithm Implementation using Patchwork Technique," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-5.

[15] Y. Yang, M. Lei, M. Cheng, B. Liu, G. Lin and D. Xiao, "An audio zero-watermark scheme based on energy comparing," in China Communications, vol. 11, no. 7, pp. 110-116, July 2014.

[16] W. Weina, "Digital audio blind watermarking algorithm based on audio characteristic and scrambling encryption," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2017, pp. 1195-1199.

[17] Allwinnaldo, G. Budiman, L. Novamizanti, R. N. Alief and M. R. R. Ansori, "QIM-based Audio Watermarking using Polar-based Singular Value in DCT Domain," 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2019, pp. 216-221.

[18] A. A. Attari and A. A. B. Shirazi, "Robust and Transparent Audio Watermarking based on Spread Spectrum in Wavelet Domain," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), 2019, pp. 366-370.

[19] G. Budiman, A. B. Suksmono, and D. Danudirdjo, Blind audio watermarking by hybrid method with high imperceptibility, robustness and capacity, Ph.D. dissertation, Bandung Institute of Technology, 2021.

[20] ITU, R, "Method for objective measurements of perceived audio quality," ITU-R Recommendation BS, 2001, 1387.

[21] Y. Itoh, K. Tajima and N. Kuwabara, "Measurement of subjective communication quality for optical mobile communication systems by using mean opinion score," 11th IEEE International Symposium on Personal Indoor and Mobile Radio Communications. PIMRC 2000. Proceedings (Cat. No.00TH8525), 2000, pp. 1330-1334 vol.2.