

Modified DCT-based Audio Watermarking Optimization using Genetics Algorithm

Ledy Novamizanti*, Gelar Budiman, Irma Safitri

Electrical Engineering Faculty, Telkom University, Bandung, Indonesia

*Corresponding author, e-mail: ledyaldn@telkomuniversity.ac.id

Abstract

Ease process digital data information exchange impact on the increase in cases of copyright infringement. Audio watermarking is one solution in providing protection for the owner of the work. This research aims to optimize the insertion parameters on Modified Discrete Cosine Transform (M-DCT) based audio watermarking using a genetic algorithm, to produce better audio resistance. MDCT is applied after reading host audio, then embedding in MDCT domain is applied by Quantization Index Modulation (QIM) technique. Insertion within the MDCT domain is capable of generating a high imperceptible watermarked audio due to its overlapping frame system. The system is optimized using genetic algorithms to improve the value of imperceptibility and robustness in audio watermarking. In this research, the average SNR reaches 20 dB, and ODG reaches -0.062. The subjective quality testing on the system obtains an average MOS of 4.22 out of five songs tested. In addition, the system is able to withstand several attacks. The use of M-DCT in audio watermarking is capable of producing excellent imperceptibility and better watermark robustness.

Keywords: audio watermarking, MDCT, genetic algorithm, imperceptibility, robustness

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

The presence of the Internet makes the exchange of information can be done very easily by anyone and anywhere, especially in publishing digital content such as music, songs, and other audio files. The number of copyright infringement cases that occur so we need a method to protect the copyright of the owner of the work. One solution is to use a watermarking method. Audio watermarking is a method of embedding a secret information into a audio host file, provided it does not damage audio files and can not be perceived by the human sense of hearing (inaudible) [1]. The embedded secret data must be resistant to various attacks and can be extracted again. Imperceptibility is a property in which the existence of a secret message can not be recognized only with the human senses, while resistance is the level of watermark technique resistance to destructive action. Many research methods can produce high levels of imperfections but are still susceptible to damage, and vice versa.

Over the last decade, MDCT has emerged as the most effective transform for audio coding due to its time domain alias cancellation and energy compaction property [2-5]. The MDCT is a derivative of DCT level IV, where this transformation is devoted to overlapping data blocks [6]. Using overlapping mechanisms, IMDCT can accurately reconstruct the original signal and avoid data aliasing at the audio frame boundaries [7]. Jianghua Li [8] discussed several basic factors that work in all DCT-based digital watermarking algorithms. In [9], watermark embedding in M-DCT domain can solve the problem in other transformation, that is can minimize artifact (distortion) that arise between data block due to frame formation. In [10], MDCT domain for watermarking improved the watermark payload or capacity.

To improve the performance of MDCT in audio watermarking, we use genetic algorithm (GA) to get the optimal embedding parameter and raise the overall performance. GA algorithms produce a sequence or series of populations using selection, crossover, and mutation mechanisms as a search mechanism [11, 12]. GA is capable to manage the trade off between robustness, imperceptibility and payload. Watermarking optimizing method with GA was found in several papers, such as in [13-16]. In [15, 16], we used GA to optimize several embedding parameters in image watermarking and the result was a significant improvement either in imperceptibility or robustness.

Quantization Index Modulation (QIM) is an embedding method for audio watermarking used in this paper. This kind of technique is popular embedding method in watermarking because of its capability on improving robustness at watermark while maintaining imperceptibility on watermarked audio. Chen and Wornell in [17] described QIM theoretically and applied it firstly in audio watermarking, after then their paper was becoming the main reference for audio watermarking on QIM embedding method by many researchers. QIM is frequently used in the frequency domain of the host audio as described in [18-20].

In this paper, we develop audio watermarking in MDCT domain of host audio with QIM embedding method and GA optimizes the embedding parameter to improve the watermarking performance. After reading host audio, MDCT transforms the audio signal into double multi-band signal in frequency domain. This multi-band signal is quite similar but unequal with previous published papers in [21] and [22]. In that study, the multi-band was generated according to the watermark, but in this paper, multi-band is the transformation result of host audio by MDCT. The watermark is then embedded in MDCT domain of audio signal using Direct Sequence Spread Spectrum (DSSS) based technique. DSSS is a transformation technique that serves to spread a signal input in time domain [23]. In the implementation the signal is multiplied 'directly' by a random number of random numbers called pseudo noise sequence (PN Sequence). In [24-28], SS based watermarking schemes can improved the watermark payload or capacity. Next, QIM is applied on each watermark bits at the certain subband of all time slots where GA optimize the position of subband to obtain highest imperceptibility and robustness. After that, frequency domain signal is transformed again into time domain by inverse MDCT obtaining watermarked audio. GA optimizes the embedding and extraction process on certain attack with decided target called fitness function (FF). FF consists of several target/output parameter which will be optimized. In this paper, the output optimize parameter are BER (robustness) and ODG (imperceptibility).

The structure of this paper consists of several sections. Section 1 describes introduction, section 2 describes research method and basic formulation of audio watermarking method, section 3 describes audio watermarking system with embedding and extraction procedures, section 4 describes the performance of the method, and finally section 5 describes the audio watermarking method conclusion.

2. Audio Watermarking System

In this study, we designed an audio watermarking system using MDCT method and performed system optimization with genetic algorithm. Scheme of the system model proposed in this study is generally illustrated in Figure 1. The first process is to insert a message into the host audio using the MDCT method [6]. Then the audio host is tested with several attacks. Optimization is done on audio that has the most damage when attacked. The system evaluated by the genetic algorithm as a whole to get the most optimal insertion parameters.

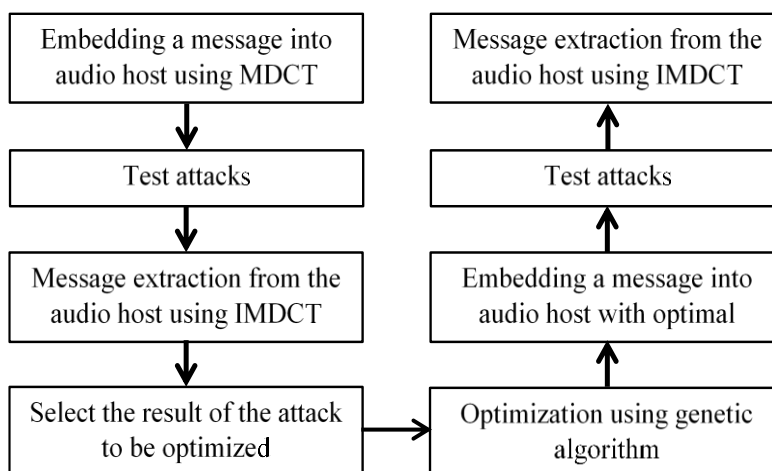


Figure 1. Scheme of the proposed method

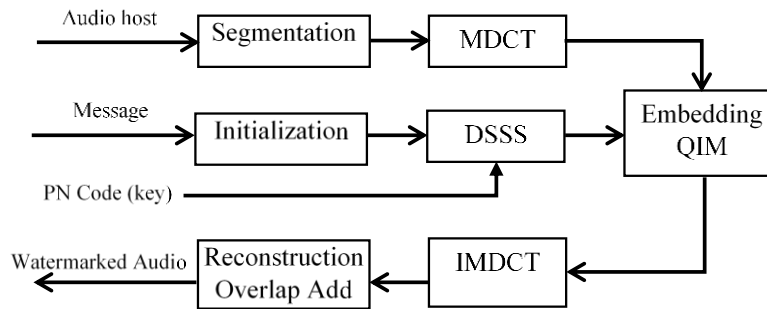


Figure 2. Block diagram of embedding process using MDCT

2.1. Embedding System Design

The MDCT-based embedding process is described in Figure 2. In the segmentation process, the original audio host in the form of a single column matrix is segmented into a frame with 50% overlap. The purpose of the segmentation process is that the latest frame contains half of the contents of the previous frame. The segmented audio is then transformed into the frequency domain using MDCT. The MDCT processes a signal of length $2N$ to form a number of N coefficients [6]. If the length of the frame is 128, and the length is considered to be a $2N$ value, the resulting MDCT coefficient is half of the frame length of 64. This applies equally to other frame lengths. The general formula of MDCT is represented by the following equation [6]:

$$X(k) = \sum_{n=0}^{2N-1} x(n) \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right] \quad (1)$$

$$x(n) = h(n) \times a(n) \quad (2)$$

$$h(n) = \sin \left[h(n) \times \left(n + \frac{1}{2} \right) / 2N \right] \quad (3)$$

Here, $k = 0, 1, \dots, N - 1$, $X(k)$ is the coefficient of MDCT, $a(n)$ is an input signal with the number of $2N$ samples, and $h(n)$ is a sinus windowing function. Windowing on MDCT serves to reduce the impact of discontinuities due to frame cutting. Calculations with combined MDCT formulas and sinusoidal window functions allow signals to be changed from time domain to frequency domain and can be reversed appropriately. Signals that have been generated with MDCT can be returned to the time domain by inverse MDCT (IMDCT) represented by the following equation [5, 6]:

$$Y(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right] \quad (4)$$

with $n = 0, 1, \dots, 2N - 1$.

Before entering the embedding stage, the size of the message is adjusted first by converting the binary image matrix from 32×32 to a line of 1×1024 . Then, PN generator will produce PN code consisting of random number is 0 and 1, with number 64 bit. The PN code is used as a key to provide more protection against secret messages [29]. The resulting PN code is then embedded or spread into messages using the Direct Sequence Spread Spectrum (DSSS) method. In this process every 1 bit PN code will be XOR to 16 bit value of the image. In this paper, DSSS process is formulated as follows:

$$P = \begin{cases} 1, & \text{if message bit} = \text{PN code bit} \\ 0, & \text{if message bit} \neq \text{PN code bit} \end{cases} \quad (5)$$

Here, P is the result of a message that has been multiplied by the PN Code.

Quantization Index Modulation (QIM) is classified as a host-interference rejection technique, which does not require host signals in the decoder [15]. The insertion is performed on frames that have an average coefficient above the threshold value. It aims to avoid

messages embedded on the MDCT coefficient value that is too small, which can cause messages easily lost when exposed to attacks. Next, the message will be embedded with QIM technique. First step is initialized by determining the quantization step value (Δ) based on the following equation:

$$\Delta = \frac{1}{2^{(nbit-1)}} \tag{6}$$

where nbit is a quantization bit in QIM. In this study, the value of nbit be varied from 1 to 10, which determined by the genetic algorithm. Furthermore, the embedding process with QIM is done by the following formula [15]:

$$\widehat{F}(n) = \begin{cases} A_k, & \text{if } P_i = 1 \text{ and } \arg \min |F(n) - A_k| \\ B_k, & \text{if } P_i = 0 \text{ and } \arg \min |F(n) - B_k| \end{cases} \tag{7}$$

$$A_k = \left(2k + \frac{1}{2}\right)\Delta, \quad B_k = \left(2k - \frac{1}{2}\right)\Delta, \quad k = 0, \pm 1, \pm 2, \tag{8}$$

Here, $\widehat{F}(n)$ is the MDCT coefficients from the quantization of $F(n)$, and P_i is the message that will be embedded. Every single bit message is inserted in one frame of the MDCT, and repetition of the message as many as the number of frames that meet the threshold. MDCT coefficient inserted message must then be converted back to the time domain by performing an inverse MDCT. Frame reconstruction process is also done simultaneously when IMDCT process takes place. The reconstruction process is carried out by summarizing the overlapping part of the frame. After IMDCT, the audio watermark will be in the time domain, and the embedding process is complete.

2.2. Extracting System Design

At this stage, the message contained in the audio watermark is retrieved. Basically the stages in the extraction process are almost the same as the stage done in the embedding process. The process is done after the MDCT coefficient obtained. Then do the demodulation process and unlock the message. The extraction process is shown in Figure 3.

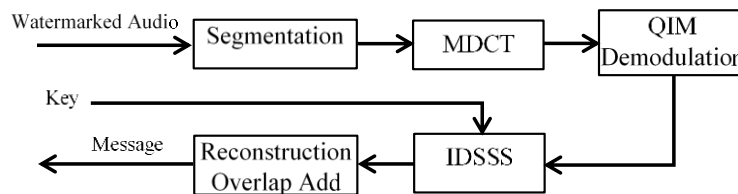


Figure 3. Block diagram of extracting process using MDCT

In the segmentation process, the watermarked audio is segmented into a 50% overlap frame. The size of each frame depends on the selected variation in the embedding process. Watermarked audio that have been shaped frames are then transformed into the frequency domain using MDCT. Then the signal is extracted with the QIM technique to retrieve the messages that have been embedded. The extraction process is carried out by the following formula [15]:

$$\widehat{P}_i = \text{mod} \left(\left\lfloor \frac{\widehat{F}(n)}{\Delta} \right\rfloor, 2 \right) \tag{9}$$

where $\widehat{F}(n)$ is the MDCT coefficient that has been embedded message, and \widehat{P}_i is the message extraction results. Because the message is embedd with repetition across all frames that meet the threshold, then the next is calculated the average of the extraction results to get the desired message. Messages that have been extracted from the audio watermark can not be read directly because the message still contains a PN code key. Therefore, at this stage the key is

unlocked by performing XOR between the PN code and the message. The value of PN code is obtained by inputting the key that has been obtained from the embedding process. After separation the key of the message, the message is still in the form of a line is to be returned to its original size and the extraction process ends.

2.3. Optimization using Genetic Algorithm

The genetic algorithm works by evaluating all parts of the watermarking process, including insertion, attack and extraction. From the optimization process, an optimal watermarking parameter will be obtained. The process of the genetic algorithm is described in Figure 4.

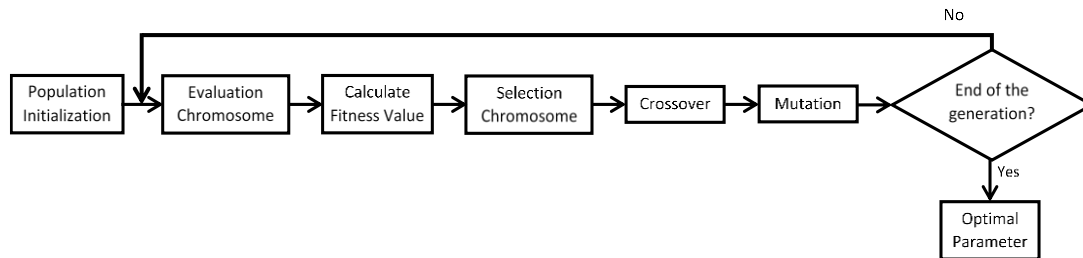


Figure 4. Flowchart of optimization using a GA [13,14]

Population initialization is the stage of chromosome formation which will be processed in genetic algorithm. To be able to form a chromosome needs to be defined multiple values, such as the number of generations and individuals that will be increased, the probability of mutation and crossover probability. In this study, there are four types of parameters to be optimized ie frame size, nbit QIM, insertion threshold and insertion position. The chromosome encoding is performed using binary encoding with the specified chromosome length being 16. The chromosome design for each parameter is given in Table 1.

Table 1. Chromosome Design

Gene Sequence in Chromosomes	Parameter Optimization	Value
Gen 1 - 2	Frame Size	128, 256, 512, 1024
Gen 3 - 10	Position the embedding	Maximum 256
Gen 11 - 14	Nbit QIM	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Gen 15 - 16	Threshold insertion Coefficient	0.00001, 0.0001, 0.001, 0.005

Before optimization, the embedding process with QIM is done with a threshold and a fixed position. But at the time of optimization, the threshold and position values will change according to their respective ranges. The insertion position of the index represents MDCT coefficients that will be embedded in message. The maximum number of embedded positions is as much as 256 grades. Selection of this parameter is expected to decrease the value of BER after the audio is attacked. In this study, evaluation of chromosome means insertion process using new parameters formed from previous process. The insertion process is carried out as described in section 2.1. Then, the watermark audio results quality is calculated by the value of SNR and ODG.

Test attack will be done on the newly formed audio watermark. Test attack on the optimization process is to test the attacks that generate BER is not 0 in the message. The types of attacks carried out, namely LPF attacks, BPF, MP3 compression, noise, and time scale modification (TSM). After an attack test, the message contained in the audio watermark will be extracted. Then the message of extraction is assessed its quality by calculating the value of BER. Fitness values are calculated to measure chromosome performance that has undergone various attack tests. The optimization process stops when a generation has reached the maximum fitness value. In this study, the fitness value is calculated based on the value of ODG and BER, with the maximum value is 1. The fitness formula used is as follows:

$$FF = 0.5 \times (1 - \overline{BER}) + 0.5 \times \left(\frac{\overline{ODG}}{4}\right) \quad (10)$$

where \overline{BER} and \overline{ODG} are the average value of BER and ODG. In the equation shows that the ODG has a greater weight than the BER is 1:1, it is to get a level of imperceptibility and robustness balanced.

The selection process is done by sorting the fitness value that has been obtained. Then taken two chromosomes that have the highest fitness value. The two chromosomes are crossed with each other, and the mutation process is performed. Termination criteria are when the chromosomes reach the highest fitness value, thus resulting optimal insertion parameters. If the termination criterion has not been reached then the system will iterate the process by inputting the new parameter. After the termination criterion is achieved, the optimal value will be stored. Furthermore, the insertion process is based on these parameters.

3. Result and Analysis

System testing is carried out in 3 stages, namely measurement of system performance before the optimization process, after optimization, and performance comparison with the previous method. Audio host file is data with *.wav format that has sample rate specification 44100 Hz, 16 bits per sample, mono, with audio duration of approximately 30 seconds. Messages to be embedded into the audio are binary images of 32×32 pixels. The assessment is conduct using 5 different types of audio as a host, namely voice, instrumental, country, jazz, and rock music.

3.1. Performance System Before Optimization Process

In this section, the effect of frame size and nbit QIM will be tested on system performance. Futhermore, an attack will be given to evaluate system performance.

a. Effect of Frame Size on SNR, ODG and BER

In this scenario, a process of embedded and extraction of QIM uses variation of frame size, with fixed parameters: Nbit 1, Threshold 0.0001, Insertion position 1 (first coefficient on each frame). The performance of ODG, SNR and BER measurements of the five tested audio are as follows:

Table 2. Effect of Frame Size on SNR, ODG, and BER

Frame Size	ODG	BER	SNR (dB)
128	-2.6178	0	0.7100
256	-1.7564	0	3.7560
512	-0.9810	0	6.9288
1024	-0.2283	0	10.2819

From Table 2, it can be seen that BER is constant 0 to all frame types. This means that messages can be extracted perfectly on each audio and each frame type. SNR and ODG value is still very low. There is a tendency that the larger the frame size the better the ODG and SNR values. The larger the frame size, the smaller the number of frames that are formed, therefore the number of embedded repetition messages will be decreased. The ratio of signal with noise will be higher, and produce better ODG. Thus, the frame size of 1024 will be used in the next process because it has the highest ODG and SNR.

b. Effect of Nbit on SNR, ODG and BER.

Nbit represent the number of quantization bits in QIM, where the nbit value will greatly affect the quantization step (Δ). In this test, QIM embedded will be performed using different nbit values, and with a frame size of 1024. The effect of nbit values on ODG, SNR and BER of the five tested audio shown in Table 3.

Table 3 shows that the larger the value of nbit the better the value of ODG and SNR obtained. The larger the value of nbit means the quantization step value (Δ) will be smaller, the signal quality will get better because the signal is quantized very close to the original signal. Thus the parameters used for the next process is the value of nbit 10 and frame size 1024.

Table 3. Effect of Nbit on SNR, ODG, and BER

Nbit	ODG	BER	SNR (dB)
1	-0.2283	0	8.4762
2	-0.1425	0	14.3261
4	-0.0795	0	25.6986
8	-0.0632	0	42.9332
10	-0.0624	0	48.7452

c. Effect of Attacks on BER.

In this section, a system of endurance testing is carried out by giving several attacks on the audio watermarking system. In the BPF attack the filter type used is the Butterworth Infinite Impulse Response (IIR). Each attack has different levels of intensity. Table 6 shows that before optimization, the watermarking audio system is still very weak against several types of attacks. This is caused by improper insertion position. In the next section, system optimization will be applied by determining the right combination of parameters with GA.

3.2. Performance System After Optimization Process

In this section, we describe the results obtained from the optimization process with genetic algorithms. Then, we analyze the results of the attack test using the optimal parameters that have been obtained. Finally, the test results is compared between before and after optimization.

a. Parameter Optimization with Genetic Algorithm

A GA is used to determine the correct combination of insertion parameters. There are four parameters to be optimized such as frame size (*nframe*), *nbit*, *position*, and *threshold*. The values for each parameter have been determined in Table 1. The genetic algorithm is performed with the following conditions: number of generations=300, number of individuals=20, crossover=0.8, mutation=0.5. To obtain the optimal parameters, testing is done with one type of attack that MP3 compression at a rate of 96 kbps, and one type of audio that is Rock.wav. MP3 compression chosen because it is a very common type of attack happens in the real world. While rock.wav is the audio that has the worst BER on the attack. Testing was carried and the result is shown in Table 4.

Table 4. Effect of Frame Size on SNR, ODG, and BER

Generation	Fitness	Frame Size	Nbit	Position	Threshold	ODG	BER	SNR (dB)
1	0.983377	512	6	97	0.0001	-0.125	0.001953	38.615
55	0.986063	512	6	225	0.0001	-0.049	0.015625	38.919
59	0.992371	512	6	229	0.0001	-0.049	0.00293	38.915
81	0.993703	512	2	229	0.0001	-0.050	0	21.047
95	0.993788	512	2	231	0.0001	-0.05	0	21.046
119	0.993841	512	4	231	0.0001	-0.049	0	32.340
144	0.995046	128	4	58	0.0001	-0.039	0	27.027

In Table 4, it can be seen that the optimal parameters obtained after entering the 144th generation with the highest fitness value is 0.995046, and BER is worth 0. BER change is quite significant when compared with MP3 compression because the insertion position of great influence on the level of resistance watermarking. Then, the value is used for the embedding of the five types of audio. The comparison of results obtained from the embedding before and after obtaining the optimal parameters is described in Table 5.

Table 5. The Effect of Optimization on System Performance

Audio	Before Optimization (Nframe = 1024, nbit = 10, position = 1, threshold = 0.0001)			After Optimization (Nframe = 128, nbit = 4, position = 58, threshold = 0.0001)		
	ODG	BER	SNR	ODG	BER	SNR
Country	-0.213	0	37.327	-0.018	0	22.093
Jazz	-0.072	0	53.395	-0.029	0	19.625
Orchestra	-0.008	0	38.793	-0.076	0	16.552
Rock	-0.003	0	41.078	-0.013	0	27.096
Voice	-0.015	0	56.875	-0.175	0	13.020
Average	-0.0624	0	45.4935	-0.0622	0	19.677

As shown in Table 5, the average SNR value decreases after optimization. Frame size is smaller enough than before, resulting in the repetition of message insertion becomes more. So the ratio between signal to noise decreases. However, the ODG values do not vary much with before optimization. When the ODG value is still above -1 and BER is 0, it means that audio still has a very good quality when heard and the message can be extracted perfectly.

b. Effect of Attacks on Optimal Parameters

The Butterworth IIR filter is used in BPF attacks. The result of attack testing after using the optimum parameter can be seen in Table 6.

Table 6. Comparative Effect of Attacks Before and After Optimization

Attack	Criteria	Average BER before optimization	Average BER after optimization
LPF	fc=3 kHz	0.2406	0.3406
	fc=6 kHz	0.05	0.0906
	fc=9 kHz	0.0375	0
BPF	100 Hz - 3 kHz	0.3187	0.1687
	100 Hz - 6 kHz	0.4375	0.0031
	100 Hz - 9 kHz	0.0969	0
	50 Hz - 6 kHz	0.47266	0.0039
	25 Hz - 6 kHz	0.3875	0.0031
Noise	0 dB	0.3594	0.4469
	10 dB	0.3281	0.2531
	20 dB	0.2312	0.075
Resampling	11 kHz	0.0656	0.025
	16 kHz	0.0281	0.1
	22 kHz	0.0031	0.1031
	24 kHz	0.1031	0.1031
Time Scale Modification	1%	0.2625	0.3531
	2%	0.2344	0.4125
	3%	0.4156	0.4094
	4%	0.4031	0.45
Linear Speed Change	1%	0.0844	0
	5%	0.0812	0
	10%	0.0875	0
Equalizer		0.0687	0.1062
Echo		0.3781	0.0562
MP3 Compression	32 kbps	0.2687	0.0094
	64 kbps	0.0937	0
	128 kbps	0.0719	0
	192 kbps	0.0812	0
Average		0.2033	0.12547

The results shows that after the optimization, the change in insertion frequency positioning makes the system highly resistant to LPF with $fc > 6$ kHz, and BPF with an frequency upper limit of ≥ 3 kHz. For compression attacks, the system is very resistant to all three types of compression rates. In addition, the system is also resistant to noise attack up to 20 dB intensity for all audio except jazz.wav. At TSM attack, BER produced is not a big change from before. It can be concluded that the system is not resistant to attacks that can modify the tempo of audio. This is because the system is very sensitive to the position of insertion. If the tempo is accelerated or slowed the insertion position can shift and cause a change of BER result.

c. Comparative Effect of Attacks Before and After Optimization

The comparison shows how the success of genetic algorithms in optimizing the audio watermarking system. Before optimization system can not tolerate any attack. However, after the optimization of the system resistant to seven attacks, namely LPF, BPF, Noise, Resampling, Linear Speed Change, Equalizer, and MP3 Compression. This is due to changes in the embedded frequency position after optimization that is no longer in the low frequency range. It can be concluded that the genetic algorithm is able to optimize the embedded parameters in the audio watermarking system based on MDCT.

d. Mean Opinion Score (MOS)

MOS is a quality assessment of audio watermarking subjectively [30]. The MOS test is done by asking 30 respondents to compare the original audio quality with audio that has been inserted message. Subjective Performance (MOS) can be seen in Table 7. The results obtained from the measurement of MOS are as follows:

Table 7. Subjective Performance (MOS)

Audio	MOS
Country	4.47
Jazz	4.43
Orchestra	3.87
Rock	4.03
Voice	4.30

Table 7 shows that there are four kinds of audio that are considered to have good quality and only one audio that is quite good is Orchestra. In the audio orchestra genre there is a voice instrument violin with dominance of high frequency. The sound of screeching and resembles the noise, so that the respondents' assessment on the audio is not good, although the value of the resulting excellent ODG. It can be said that this system has good quality in terms of subjective judgement, that is $MOS \geq 3.87$.

3.3. Performance Comparison with Previous Method

MDCT in audio watermarking has good contribution to increase the imperceptibility. In the same time, this method has good robustness to several attacks as displayed in Table 4. In the previous paper about audio watermarking with MDCT method, there were lack in reporting the performance especially in robustness. In [9] and [10], MDCT is used as audio watermarking method, but there are only imperceptibility reported instead of robustness. In Table 8, we display the performance comparison between our method with previous method on 4 type attacks, MP3 32-64 kbps, echo attack, and noise attack with 20 dB power.

Table 8. Performance Comparison

Ref	Robustness				Imperceptibility	
	MP3 32 kbps	MP3 64 kbps	Echo	Noise 20 dB	SNR (dB)	ODG
[9]	NA	NA	NA	NA	NA	-0.34
[10]	NA	NA	NA	NA	NA	-0.15
[19]	NA	2.96%	4.10%	0.92%	18.21-21.91	0.005
[20]	7.60%	0.90%	9.20%	NA	12.5	NA
[22]	47%	9%	0%	0%	8.62	-2.2
Proposed	0.94%	0%	5.62%	7.50%	19.68	-0.06

NA means not available or there is no performance reporting. MDCT with optimized parameter by Genetics Algorithm gives accepted performance in robustness with those four types of attacks and gives high performance in imperceptibility. The robustness of the method has worst result in noise attack with BER 7.5% in which it is still below 10% or acceptable robustness.

4. Conclusion

Implementation of MDCT-based audio watermarking with Genetics Algorithm optimization in this work is able to produce high audio watermarking imperceptibility where $ODG > -0.2$ with average ODG is -0.06 and with MOS is more than 4. This proposed method is robust against several attacks indicated by BER which is less than 10%, such as LPF with cut off frequency ≥ 6 kHz, BPF with cut off frequency 25/50/100 Hz – 6/9 kHz, noise 20 dB, linear speed change, echo, and MP3 attacks with rate up to 32 kbps. But, the proposed method with optimized parameters is still not robust against TSM, resampling and equalizer attacks because optimizing is done by different attack. With unoptimized parameter, the proposed method is already robust against resampling and equalizer attack.

References

- [1] Chauhan SPS, Rizvi SAM. *A survey: Digital Audio Watermarking Techniques and Applications*. International Conference on Computer and Communication Technology (ICCCT). India. 2013.
- [2] Johnson AW, Bradley AB. *Adaptive transform coding incorporating time domain aliasing cancellation*. Speech Comm. 1987; 6 (4): 299-308.
- [3] Malvar HS. *Signal Processing with Lapped Transforms*. Norwood: Artech House Publisher. 1992.

- [4] Yu R, Lin X. *A Statistics study of the MDCT coefficient distribution for audio*. IEEE International Conference on Multimedia and Expo (ICME). Taiwan. 2004: 1483-1486.
- [5] Raj VA, Dhas MDK, Gnanadurai D. *An overview of MDCT for Time Domain Aliasing Cancellation*. International Conference on Communication and Network Technologies. India. 2014: 203-207.
- [6] Shao X, Johnson SG. Type-IV DCT, DST, and MDCT algorithms with reduced numbers of arithmetic operations. *Signal Processing*. 2008; 88(6): 1313-1326.
- [7] Wang ML, Lin HX, Lee MT. *Robust audio watermarking based on MDCT coefficients*. International Conference on Genetic and Evolutionary Computing. Japan. 2012: 372-375.
- [8] Li J, Cao Q. DSDWA: A DCT-based Spatial Domain Digital Watermarking Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2015; 12(1): 693-702.
- [9] Singh SK, Singh J. Audio Watermarking Scheme in MDCT Domain. *International Journal of Electrical and Electronics Engineering (IJEET)*. 2012; 1(3): 42-47.
- [10] Pinel J, Girin L, Baras C, Parvaix M. *A high-capacity watermarking technique for audio signals based on MDCT-domain quantization*. Int. Congr. of Acoust. Australia. 2010: 1-7.
- [11] Sheikh RH, Raghuvanshi MM, Jaiswal AN. *Genetic Algorithm Based Clustering: A Survey*. First Int. Conf. Emerg. Trends Eng. Technol. India. 2008; 2(6): 314-319.
- [12] Suyanto, *Algoritma Genetika Dalam Matlab*. 2005. Yogyakarta: Andi Offset.
- [13] Sadeghzadeh M, Taherbaghal M. *A New Method for Watermarking using Genetic Algorithms*. Int. Conf. Mach. Learn. Electr. Mech. Eng. Dubai. 2014: 1-8.
- [14] Song W, Sun X, Liu C, Tang L. A New Watermarking Frame Based on the Genetic Algorithms and Wavelet Packet Decomposition. *Inf. Hiding Multimed. Signal Process*. 2015; 6(3): 613-621.
- [15] Budiman G, Novamizanti L, Iwut I. Genetics Algorithm Optimization of DWT-DCT Based Image Watermarking. *Journal of Physics: Conference Series*. 2016; 795(1): 1-8.
- [16] Iwut I, Budiman G, Novamizanti L. Optimization of Discrete Cosine Transform-Based Image Watermarking by Genetics Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2016; 4(1): 1-16.
- [17] Chen B, Wornell GW. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theory*. 2001; 47(4): 1423-1443.
- [18] Zhou Z, Zhou L. *A Novel Algorithm for Robust Audio Watermarking Based on Quantification DCT Domain*. International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Southern Yangtza University. Taiwan. 2007: 441-444.
- [19] Hu H, Chen S, Hsu L. *Incorporation of Perceptually Energy-compensated QIM into DWT-DCT Based Blind Audio Watermarking*. International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Japan. 2014: 748-752.
- [20] Khademi N, Akhaee MA, Ahadi SM, Moradi M, Kashi A. *Audio watermarking based on Quantization Index Modulation in the Frequency Domain*. IEEE Int. Conf. Signal Process. Commun. Dubai. 2007: 1127-1130.
- [21] Aparna S, Baiju PS. *Audio Watermarking Technique Using Modified Discrete Cosine Transform*. International Conference on Communication Systems and Networks (ComNet). India. 2016.
- [22] Budiman G, Suksmo AB, Danudirdjo D, Usman K, Shin DH. A modified multicarrier modulation binary data embedding in audio file. *Int. J. Electr. Eng. Informatics*. 2016; 8(4): 762-773.
- [23] Kirovski D, Malvar H. *Spread-spectrum audio watermarking: requirements applications and limitations*. *IEEE Fourth Workshop on Multimedia Signal Processing*. France. 2001: 219-224.
- [24] Kirovski D, Malvar H. *Robust spread-spectrum audio watermarking*. IEEE Int. Conf. Acoust. Speech Signal Process. USA. 2001: 3-6.
- [25] Kirovski D, Malvar HS. Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing*. 2003; 51(4): 1020-1033.
- [26] Singh J, Garg P, De AN. Audio Watermarking using Spectral Modifications. *International Journal of Electronics and Communication Engineering*. 2009; 3(11): 2207-2211.
- [27] Chen G, Yang S, Zhang Q, Du J. *Robustness Performance Analysis of Audio Watermarking Algorithm Based on Direct Sequence Spread Spectrum*. International Conference on Internet Technology and Applications. China. 2011.
- [28] Cox IJ, Kilian J, Leighton FT, Shamoon T. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*. 1997; 6: 1673-1687
- [29] Swanson MD, Zhu B, Tewfik AH, Boney L. Robust audio watermarking using perceptual masking. *IEEE Transactions on Signal Processing*. 1998; 66: 337-355.
- [30] Lin Y, Abdulla W. Objective quality measures for perceptual evaluation in digital audio watermarking. *Signal Process. IET*. 2011: 623-631.