

PERANCANGAN DAN ANALISIS AUDIO WATERMARKING BERBASIS MODIFIED DISCRETE COSINE TRANSFORM (MDCT) DENGAN METODE STATIONARY WAVELET TRANSFORM (SWT) DAN CENTROID

DESIGN AND ANALYSIS AUDIO WATERMARKING BASED ON MODIFIED DISCRETE COSINE TRANSFORM (MDCT) WITH STATIONARY WAVELET TRANSFORM (SWT) AND CENTROID

Mohammad Bisma Rezady¹, Gelar Budiman, S.T., M.T.², Irma Safitri, S.T., M.Sc.³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

¹bismarezady@gmail.com, ²gelar.budiman@gmail.com, ³irmasaf@telkomuniversity.ac.id

Abstrak

Audio watermarking merupakan solusi untuk perlindungan hak cipta yang dapat menyisipkan data atau *watermark* kedalam *host audio* yang kehadirannya tidak diketahui indra manusia. Metode yang digunakan dalam jurnal ini adalah *Modified Discrete Cosine Transform (MDCT)*, *Stationary Wavelet Transform (SWT)* dan *Centroid*. Langkah pertama proses penyisipan adalah *host audio* disegmentasi menjadi beberapa frame kemudian dilakukan proses transformasi SWT untuk membagi sinyal kedalam beberapa *sub-band*. *Sub-band* yang telah dipilih untuk dilakukan penyisipan ditransformasi dengan MDCT untuk mengurangi distorsi yang terjadi antar *frame* blok, kemudian proses *centroid* digunakan untuk mencari titik tengah frekuensi untuk lokasi penyisipan agar keluaran yang dihasilkan lebih stabil dan penyisipan *watermark* dilakukan dengan metode using *Quantization Index Modulation (QIM)*. Dengan menggunakan metode tersebut, hasil dari keluaran sistem meraih nilai SNR = 27,9387 dB, ODG = -3,7976 dan kapasitas = 21,5332 bps dengan nilai rata-rata MOS = 4,046. Selain itu sistem juga tahan terhadap berbagai jenis serangan dengan nilai BER mencapai 0 (6 dari 11 serangan).

Kata kunci : MDCT, SWT, Centroid, QIM

Abstract

Audio watermarking is a solution for protecting copyright that can hide data or digital information on host audio that mostly unknown by human sense. The method that used in this paper are Stationary Wavelet Transform (SWT) and Centroid based on Modified Discrete Cosine Transform (MDCT). The first step the host audio is segmented into several frames then transformed using SWT to get the signal into several sub-band. Next step the sub-band which selected will be processing by using MDCT to reduce blocking artifacts and centroid to get the center of spectrum for more stable result and the embedding process is using Quantization Index Modulation (QIM). By using that methods, the result can reach imperceptibility with SNR = 27, 9387 dB, ODG = -3,7976, capacity = 21,5332 bps and the average MOS = 4,046 and also high robustness that resistant to various attacks with the value of BER approaching 0 (6 out of 11 attacks).

Keywords: MDCT, SWT, Centroid, QIM

1. Pendahuluan

Watermarking adalah suatu cara untuk menyisipkan data atau informasi tertentu ke dalam suatu data digital namun hasil dari *watermarking* tidak dirasakan oleh indera manusia (indera penglihatan atau indera pendengaran). Dalam *audio watermarking*, data yang akan disisipkan disebut *watermark* sedangkan konten audio yang akan dilakukan penyisipan *watermark* disebut *host audio*. Penyisipan informasi kedalam data digital dilakukan dengan memperhatikan berbagai parameter agar tidak merusak kualitas data yang disisipi. Pada penelitian [1] dengan menggabungkan metode *Discrete Cosine Transform (DCT)* dan *Centroid* didapatkan hasil ekstraksi yang sempurna. Namun dengan ditambahkan beberapa serangan yang dilakukan, terlihat bahwa pada serangan MP3 (48kbps) menunjukkan hasil nilai ekstraksi terendah. Sedangkan pada penelitian [2] menunjukkan dengan menggabungkan metode *Discrete Cosine Transform (DCT)*, *Discrete Wavelet Transform (DWT)* dan *Centroid* didapatkan hasil *Objective Different Grade (ODG)* yang mendekati 0 namun *Signal to Noise Ratio (SNR)* yang semakin menurun dengan bertambahnya nilai kuantisasi Δ . Penggunaan MDCT dalam penelitian ini bertujuan untuk mengurangi distorsi yang terjadi antar blok dengan sistem *overlapping* 50% dimana dinilai mampu merekonstruksi sinyal dengan sempurna seperti dalam penjelasan penelitian [3]. Metode SWT seperti yang dijelaskan dalam penelitian [4] bertujuan untuk menyempurnakan translasi-invarian dari DWT dengan cara meniadakan *downsampling* pada setiap level dekomposisi, sehingga masukan dan keluaran SWT bernilai sama walaupun telah melewati proses dekomposisi. Dan metode *centroid* pada penelitian [1] bertujuan untuk mencari nilai amplitudo titik pusat sinyal untuk penentuan penyisipan *watermark* agar keluaran yang dihasilkan lebih stabil.

2. Dasar Teori

2.1 Modified Discrete Cosine Transform (MDCT)

MDCT merupakan transformasi yang digunakan untuk merubah sinyal dari domain waktu kedalam domain frekuensi. Basis dari transformasi MDCT yakni adanya sistem *overlapping* 50% antara batas *frame*. Sebuah *frame* MDCT memiliki panjang N dengan panjang *frame* masukan sebesar $2N$. Persamaan untuk MDCT ditunjukkan sebagai berikut [5] :

$$X(k) = \sum_{n=0}^{2N-1} x(n) \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right] \quad (2.1)$$

$$x(n) = h(n) x a(n) \quad (2.2)$$

$$h(n) = \sin \left[\frac{\pi}{2N} \left(n + \frac{1}{2} \right) \right] \quad (2.3)$$

Dimana :

$k = 0, 1, \dots, N-1$

$a(n)$ = sinyal input dengan sampel $2N$

$h(n)$ = fungsi *windowing sinus*

$X(k)$ = koefisien keluaran MDCT

Kemudian setelah dilakukan proses MDCT, sinyal harus dikembalikan lagi kedalam domain waktu menggunakan persamaan *Invers MDCT* (IMDCT) [5] :

$$X(n) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right] \quad (2.4)$$

2.2 Stationary Wavelet Transform (SWT)

Stationary Wavelet Transform merupakan transformasi *wavelet* yang melakukan proses dekomposisi frekuensi kedalam frekuensi tinggi dan frekuensi rendah. SWT dirancang untuk menyempurnakan translasi-invarian pada DWT dengan menghilangkan *downsampling* pada DWT, sehingga SWT menggunakan *upsampling* pada setiap level dekomposisinya [6]

Sampel sinyal input yang diproses dalam SWT akan melewati respon *impuls Low Pass Filter* (LPF) dan *High Pass Filter* (HPF) dan keduanya dilakukan perhitungan konvolusi sehingga didapatkan koefisien LPF dan HPF. Berikut merupakan persamaan konvolusi pada SWT [7] :

$$y[n] = (x * g)[n] = \sum_{k=0}^{N-1} x[k]g[n-k]; -\infty < k < \infty \quad (2.5)$$

2.3 Centroid

Centroid merupakan metode yang digunakan untuk menentukan titik tengah amplituda suatu sinyal dan memiliki kelebihan terhadap ketahanan sinyal yang akan dilakukan beberapa teknik pemrosesan sinyal digital maupun berbagai jenis serangan agar memiliki hasil yang lebih stabil. Frekuensi *centroid* dari *audio frame* dihitung menggunakan persamaan berikut [8] :

$$SC_n = \frac{\sum_{k=1}^K k A_n[k]}{\sum_{k=1}^K A_n[k]} \quad (2.6)$$

Dimana :

k = indeks matriks

K = banyak sampel dalam satu *frame*

$A_n[k]$ = nilai sampel sinyal dalam *frame*

2.4 Quantization Index Modulation (QIM)

Quantization Index Modulation merupakan salah satu metode penyisipan yang sering digunakan dalam *audio watermarking*. Dalam penggunaannya QIM dapat diaplikasikan pada domain waktu maupun domain frekuensi berikut merupakan persamaan dalam QIM [9] :

$$F'(0) = A_k, \text{ if } w = 1 \text{ dan } \arg \min |F(0) - A_k| \tag{2.7}$$

$$F'(0) = B_k, \text{ if } w = 0 \text{ dan } \arg \min |F(0) - B_k| \tag{2.8}$$

$$A_k = \Delta \left(2 + \frac{1}{2} \right) \tag{2.9}$$

$$B_k = \Delta \left(2 - \frac{1}{2} \right) \tag{2.10}$$

$$\Delta = \frac{1}{2^{nbit-1}} \tag{2.11}$$

Dimana :

$k = 0, \pm 1, \pm 2, \dots$

$F(0)$ = amplituda sinyal audio sebelum proses kuantisasi

$F'(0)$ = amplituda sinyal audio setelah proses kuantisasi

w = bit *watermark*

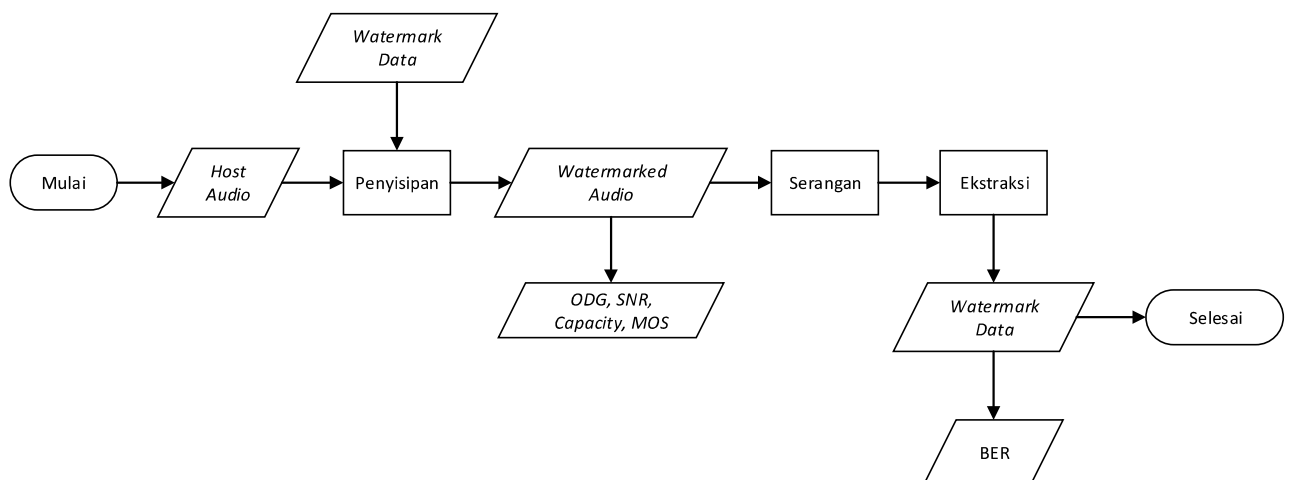
$nbit$ = jumlah bit kuantisasi

Bit *watermark* akan dilakukan ekstraksi kembali setelah disisipkan kedalam *host audio* dengan menggunakan persamaan ekstraksi QIM sebagai berikut [6] :

$$\tilde{v}(k) = \text{mod} \left(\text{ceil} \left(\frac{F'(0)}{\Delta} \right), 2 \right) \tag{2.12}$$

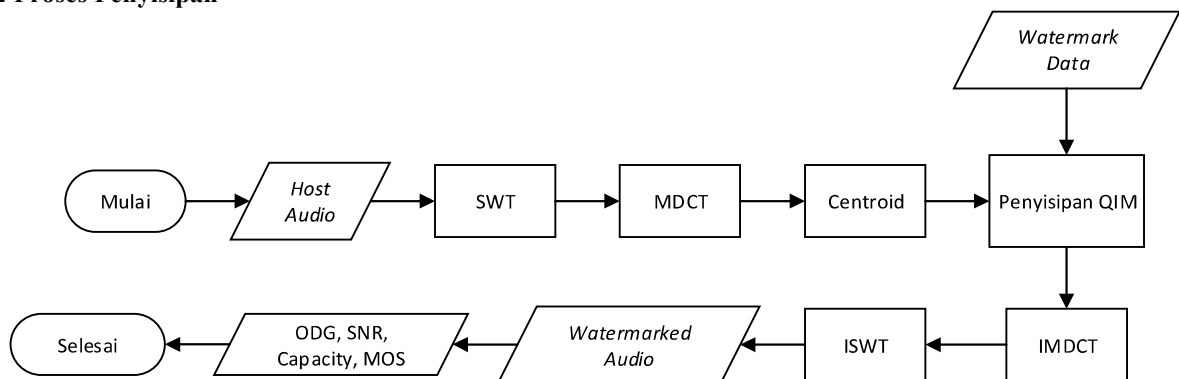
3. Perancangan Sistem

3.1 Skema Umum Sistem



Gambar 3.1 Skema Umum Sistem *Audio Watermarking*

3.2 Proses Penyisipan



Gambar 3.2 Proses Penyisipan *Watermark*

Proses penyisipan :

1. Melakukan proses *reshaping* yaitu mengubah citra biner 2 dimensi menjadi 1 dimensi.
2. Menghitung bit/frame yang akan disisipkan dengan persamaan :

$$L = \frac{\text{Panjang Frame} \times \text{Panjang Kolom Penyisipan}}{\text{Panjang Baris MDCT}} \quad (3.1)$$

3. Menentukan batas *host* minimal yang digunakan

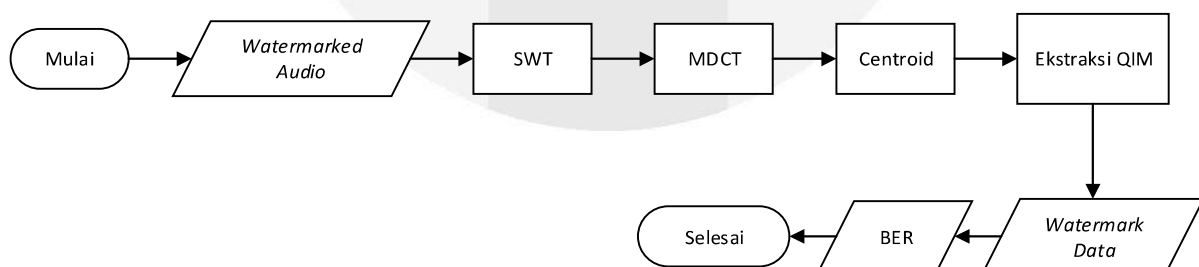
$$\text{hostmin} = \frac{\text{Panjang Frame} \times \text{Ukuran Watermark}}{L} \quad (3.2)$$

4. Membagi *host audio* yang digunakan menjadi beberapa *frame* untuk proses berikutnya.

$$\text{Jumlah frame} = \frac{\text{sampel total}}{\text{panjang sampel wavelet}} \quad (3.3)$$

5. Memproses setiap *frame* dengan SWT untuk memisahkan antara frekuensi rendah dan tinggi dengan melakukan proses konvolusi sampel sinyal input dengan respon impuls filter menggunakan persamaan (2.5)
6. Keluaran dari SWT dipilih *sub-band* dengan frekuensi rendah yang akan dilakukan proses segmentasi *frame* dengan 50% *overlapping*. Keluaran dari proses segmentasi 50% *overlapping* menjadi matriks 2 dimensi yang sebelumnya hanya 1 dimensi. Parameter input panjang baris MDCT menentukan banyaknya baris dalam 1 *frame*.
7. Kemudian keluaran dari proses segmentasi 50% *overlapping* dilakukan transformasi menggunakan MDCT dengan persamaan (2.1). MDCT akan memproses sinyal dengan panjang 2N kedalam N koefisien.
8. Keluaran dari MDCT akan dilakukan proses *centroid* menggunakan persamaan (2.6) yakni dengan menentukan nilai titik tengah amplituda sinyal untuk proses penempatan penyisipan *watermark*. Peran dari parameter input panjang kolom penyisipan akan membagi keluaran MDCT menjadi beberapa kolom untuk penyisipan setiap satu bit *watermark*.
9. Selanjutnya dilakukan proses penyisipan QIM dengan menggunakan persamaan (2.11) untuk persamaan (2.7) dan (2.9) jika bit *watermark* bernilai 1 dan persamaan (2.8) dan (2.10) untuk bit *watermark* bernilai 0. Setiap satu bit *watermark* disisipkan sesuai dengan titik tengah amplituda sinyal yang dihasilkan pada proses *centroid*.
10. Setelah proses penyisipan selesai dilakukan *invers* MDCT (IMDCT) dengan proses rekonstruksi *frame* dengan sistem *overlap-add* menggunakan persamaan (2.4). Selain itu proses IMDCT digunakan untuk mengembalikan sinyal kedalam domain waktu.
11. Keluaran dari IMDCT akan dilakukan *invers* SWT (ISWT) dengan menggabungkan bagian *sub-band* frekuensi tinggi yang tidak dipilih dalam proses penyisipan.
12. Hitung SNR, ODG, kapasitas dan survey MOS

3.3 Proses Ekstraksi



Gambar 3.2 Proses Ekstraksi *Watermark*

Proses ekstraksi pada dasarnya hampir sama dengan proses penyisipan, hanya saja dalam ekstraksi tidak diperlukan fungsi *invers* dan fungsi ekstraksi QIM yang digunakan adalah persamaan (2.12)

4. Analisis

4.1 Parameter Sebelum Dilakukan Serangan dan Optimasi

Terdapat 6 jenis parameter awal yang akan dilakukan pengujian terhadap 5 jenis *host audio* yang digunakan diantaranya : percakapan.wav, piano.wav, gitar.wav, drums.wav dan bass.wav Parameter yang akan digunakan dalam pengujian awal terdapat dalam tabel 4.1

Tabel 4.1 Parameter Sebelum Dilakukan Serangan dan Optimasi

Panjang Frame	Panjang Kolom Penyisipan	Panjang Baris MDCT	Jumlah Bit Kuantisasi	Threshold	Level Dekomposisi
65536	32	128	4	0,001	1

4.2 Parameter Setelah Dilakukan Serangan

Parameter dalam tabel 4.1 digunakan untuk proses pengujian terhadap semua jenis *host audio* dengan dilakukan beberapa jenis serangan (*low pass filter, band pass filter, noise, resampling time scale modification, linear speed change, pitch shifting, equalizer, echo, mp3 dan mp4*). Hasil rata-rata BER dari semua jenis serangan menggunakan parameter tabel 4.1 terdapat dalam tabel berikut :

Tabel 4.2 Parameter Setelah Dilakukan Serangan

Host Audio	percakapan.wav	drums.wav	piano.wav	gitar.wav	bass.wav
Rata-rata BER	0,2539	0,3169	0,4382	0,2689	0,3230

Dalam tabel 4.2, terlihat bahwa keluaran sistem setelah dilakukan serangan memiliki rata-rata BER >10% yang artinya sistem masih belum cukup tahan terhadap berbagai jenis serangan. Oleh karena itu dalam sub-bab berikutnya akan dilakukan optimasi untuk meningkatkan ketahanan *watermark* terhadap berbagai jenis serangan.

4.3 Parameter Optimasi

Setelah dilakukan serangan pada bab 4.2 dapat disimpulkan bahwa sistem *audio watermarking* tidak tahan terhadap berbagai serangan. Dalam bab ini akan dilakukan optimasi yang sebelumnya dilakukan pemilihan 5 jenis *host audio* dan jenis serangan tertentu. Pemilihan ini dilakukan dengan kriteria $0.2 < BER < 0.4$ yang masih memungkinkan untuk dilakukan optimasi. Berikut merupakan hasil pemilihan parameter yang akan dilakukan proses optimasi :

Tabel 4.3 Jenis Host Audio dan Serangan Yang Dioptimasi

Jenis Serangan	Host Audio	BER
<i>LPF Frequency Cut-Off 3000 Hz</i>	percakapan.wav	0,3438
<i>BPF Frequency Cut-Off 100-6000 Hz</i>	drums.wav	0,3945
<i>Resampling 16000 Hz</i>	piano.wav	0,4063
<i>Time Scale Modification 0.99</i>	gitar.wav	0,4063
<i>Compression MP3 32k</i>	bass.wav	0,3945

Setelah dilakukan pemilihan sesuai tabel 4.3 maka setiap *host audio* dengan jenis serangan dilakukan optimasi dengan mengubah parameter seperti panjang *frame*, panjang kolom penyisipan, panjang baris MDCT, jumlah bit kuantisasi, *threshold* dan level dekomposisi. Optimasi dilakukan hingga mendapatkan nilai BER <0,1 dan SNR >20dB. Kemudian hasil dari optimasi setiap jenis *host audio* dan serangan tersebut akan dilakukan pengujian dengan semua serangan sehingga diperoleh hasil rata-rata BER seperti dalam tabel dibawah ini :

Tabel 4.4 Rata-rata BER Semua Serangan Setelah Optimasi



















Parameter	1	2	3	4	5
Host Audio	percakapan.wav	drums.wav	piano.wav	gitar.wav	bass.wav
Rata-rata BER	0,2890	0,2863	0,2550	0,3319	0,2462














Berdasarkan tabel 4.4 parameter optimal 5 memiliki nilai rata-rata BER terkecil dengan nilai 0,2462. Namun untuk parameter optimal 5 memiliki nilai *imperceptibility* yang kurang baik, sehingga dipilih parameter optimal 3 dengan nilai rata-rata BER = 0,2550 sebagai gantinya dengan pertimbangan memiliki nilai rata-rata BER yang tidak jauh dari parameter optimal 5 namun dengan tingkat *imperceptibility* yang baik untuk semua jenis *host audio*.

Tabel 4.5 Parameter Terpilih Hasil Optimasi

Length of frame	Insertion column length	Frame MDCT	Number of quantization bits	Threshold	SWT level decomposition
65536	64	64	1	0,001	4

Tabel 4.6 Hasil Optimasi Seluruh Serangan

Jenis Serangan	BER	Hasil Ekstraksi
<i>LPF Frequency Cut-Off 3000 Hz</i>	0,0039	
<i>LPF Frequency Cut-Off 6000 Hz</i>	0	
<i>LPF Frequency Cut-Off 9000 Hz</i>	0	
<i>BPF Frequency Cut-Off 100-6000 Hz</i>	0,0234	
<i>BPF Frequency Cut-Off 50-6000 Hz</i>	0,0117	
<i>BPF Frequency Cut-Off 25-6000 Hz</i>	0,0039	
<i>Noise 0 dB</i>	0,4766	
<i>Noise 10 dB</i>	0,5234	
<i>Noise 20 dB</i>	0,2656	
<i>Resampling 22050 Hz</i>	0	
<i>Resampling 11025 Hz</i>	0	
<i>Resampling 16000 Hz</i>	0	
<i>Times Scale Modification 0.99</i>	0,4258	
<i>Times Scale Modification 0.98</i>	0,3516	
<i>Times Scale Modification 0.96</i>	0,4727	
<i>Linear Speed Change 0.99</i>	0,0078	
<i>Linear Speed Change 0.98</i>	0,0078	
<i>Linear Speed Change 0.96</i>	0,0078	

<i>Pitch Shifting 0.99</i>	0,4688	
<i>Pitch Shifting 0.98</i>	0,5273	
<i>Pitch Shifting 0.96</i>	0,4766	
<i>Equalizer</i>	0,2852	
<i>Echo</i>	0,5156	
<i>Compression MP3 32k</i>	0,0703	
<i>Compression MP3 64k</i>	0,0430	
<i>Compression MP3 128k</i>	0,0313	
<i>Compression MP3 192k</i>	0,0039	
<i>Compression MP4 32k</i>	0,0391	
<i>Compression MP4 64k</i>	0,0391	
<i>Compression MP4 128k</i>	0,0391	
<i>Compression MP4 192k</i>	0,0391	

Pada tabel 4.6 keluaran sistem yang dihasilkan cukup baik karena 6 dari 11 jenis serangan memiliki hasil ekstraksi *watermark* dengan nilai BER <0,1. Namun untuk jenis serangan *noise*, *pitch shifting*, *time scale modification*, *equalizer* and *echo* masih belum cukup tahan karena memiliki nilai BER >26% yang artinya cukup besar. Selain itu nilai SNR = 27,9387 dB dan nilai kapasitas = 21,5332 bit/s.

Selain diuji dengan pengujian objektif, sistem juga dilakukan pengujian secara subjektif dengan memperdengarkan hasil dari parameter optimasi yang digunakan untuk ke 5 jenis *host audio* kepada 32 responden. Rata-rata MOS yang diperoleh sebesar 4,046.

Tabel 4. Mean Opinion Score

<i>Host Audio</i>	MOS	ODG	SNR (dB)
percakapan.wav	3,97	-3,3931	21,1809
piano.wav	4,07	-3,7976	27,9387
gitar.wav	3,97	-2,3731	19,5237
bass.wav	4,03	-2,0204	29,4630
drums.wav	4,2	-1,4027	26,4017

5. Kesimpulan

Parameter optimal yang tahan terhadap berbagai serangan yaitu panjang frame = 65536, jumlah bit kuantisasi = 1, level dekomposisi = 4, panjang kolom penyisipan = 64, panjang baris MDCT = 64 dan *threshold* = 0,001. Namun parameter optimal hasil optimasi yang didapatkan untuk satu jenis audio, belum tentu menjadi parameter terbaik untuk jenis audio lain. Sistem *audio watermarking* yang dirancang tahan terhadap berbagai serangan dan memiliki nilai BER mendekati 0, namun untuk serangan *Noise*, *Pitch Shifting*, *Time Scale Modification*, *Equalizer* dan *Echo* memiliki nilai BER >26% yang artinya sistem *audio watermarking* yang digunakan masih belum tahan terhadap serangan tersebut. Sistem *audio watermarking* yang digunakan memiliki nilai SNR = 27,9387 dB, ODG = -3,7976 kapasitas = 21,5332 bit/s dan nilai rata-rata MOS = 4,046.

Referensi

- [1] M. Fan and H. Wang, "Centroid-based robust audio watermarking scheme," *ICALIP 2008 - 2008 Int. Conf. Audio, Lang. Image Process. Proc.*, no. 1, pp. 476–479, 2008.
- [2] W. Hongxia and F. A. N. Mingquan, "Centroid-based semi-fragile audio watermarking in hybrid domain," *Sci. China Press Springer-Verlag Berlin Heidelb.*, vol. 53, no. 3, pp. 619–633, 2010.
- [3] C. Engineering, "Audio Watermarking Technique using Modified Discrete Cosine Transform," no. July, pp. 227–230, 2016.
- [4] A. Al Jumah, "Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Techniques," vol. 2013, no. February, pp. 33–41, 2013.
- [5] X. Shao and S. G. J. Ñ, "Type-IV DCT , DST , and MDCT algorithms with reduced numbers of arithmetic operations," vol. 88, pp. 1313–1326, 2008.
- [6] C. U. I. Delong, L. Qirui, Y. U. Guilan, and X. Jianbin, "Content Based Audio Watermarking Method To Resist De-Synchronization Attacks," *ICINS 2014 Int. Conf. Inormation Netw. Secur.*, pp. 0–5, 2014.
- [7] W. Zeng, "A Novel Audio Watermarking Algorithm based on Chrip Signal and Discrete Wavelet Transform," Jiangxi Electric Power Research Institute, Nanchang, Jiangxi, China.
- [8] W. Chu, "A Noise-Robust FFT-Based Auditory Spectrum With Application in Audio Classification," *Language (Baltim)*, vol. 16, no. 1, pp. 137–150, 2008.
- [9] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

