A Modified Multicarrier Modulation Binary Data Embedding in Audio File

Gelar Budiman[1], Andriyan B. Suksmono[2], Donny Danudirdjo[3],
Koredianto Usman[4], and Dong Hwan Shin[5]

Electrical Engineering Faculty, Telkom University[1,4]
Graduate School of Electrical and Informatics Engineering Bandung Technology Institute[2,3]
Markany Inc.[5],
gelarbudiman@telkomuniversity.ac.id[1], suksmono@stei.itb.ac.id[2], donny@stei.itb.ac.id[3],
koredianto.usman@gmail.com[4], dhshin@markany.com[5]

Abstract: Information hiding of data in an audio file is an important thing that the media can be recognized by its ownership. The hidden information would be the important information which describes the copyright of the audio file. In this research, the binary data is inserted or hidden into the audio file by multicarrier modulation technique data hiding. The encoded binary data is modulated by multicarrier frequencies before embedding into the host audio. Data hiding capacity with this technique can achieve up to 40 bits per second in file mode with good audio imperceptibility. It also achieves perfect robustness for type of attack noise addition, linear speed change, multi band equalization, and echo addition. And this technique is still acceptable for type of attack, such as resampling attack, MP3 compression attack and filtering attack for cut off frequency more than 10 kHz regarding to the BER less than 10%.

Keywords: Modified multicarrier modulation, copyright, binary data, extraction, insertion, imperceptibility, attack

1. Introduction

Multicarrier modulation is a technique to modulate a baseband signal into several subcarrier at different frequencies in the same time. Prasad [1] said that OFDM basic principle is to split datastream into lower datastream transmitting simultaneously over several subcarriers. This principle is also called multicarrier modulation. The research about multicarrier modulation technique has been previously published on many papers and books. Nicola Marchetti [2] said that OFDM promises high data rate capability and uses available spectrum efficiently, which is a main reason OFDM was accepted by present telecommunication generation system.

Data hiding by the OFDM method was already presented by Somnath [3]. His research is about hiding the grayscale image in OFDM signal where the image is read pixel by pixel and then converted the binary data into a complex number format and inserted the number via Quantized Index Modulation. In [4] Amirtharajan presented the paper describing the multicarrier steganography. He presented the image hiding by Spread Spectrum method combined with the Modulated Multicarrier signal in orthogonal radio frequency, but the robustness and imperceptibility were not explained. Shishkin [5] presented audio watermarking for Electronic Radiotelephone Identification in OFDM-based. He used Quantization Index Modulation as embedding method in frequency domain, but he didn't describe the imperceptibility and robustness. Garcia-Hernandez [6] proposed a high payload data-hiding scheme for audio signals in OFDM-based. The method which he used was to change the phase component of audio signal via a reduced-arc of M-order Phase Shift Keying (MPSK). It was modulated on selected frequency of the audio signals, however he only described the imperceptibility and capacity. The robustness was not described clearly.

In this paper, we propose audio watermarking by multicarrier modulation method which the subcarrier signal are not orthogonal. This multicarrier watermarking is a watermarking process in which the hidden data is modulated by multicarrier by many frequencies from 0 to 22 kHz. The process of watermarking is not only multicarrier modulation of the hidden data, but also

Received: January 18th, 2016. Accepted: December 19th, 2016
DOI: 10.15676/ijeei.2016.8.4.5

there are several subprocessing before modulation processing and embedding the data. The imperceptibility, capacity, and robustness is described in this proposed method. The topic of this paper is modified and improved version of our previous paper in [7].

This paper is organized as follow: section 2 describes the watermarking model of multicarrier audio watermarking, section 3 presents multicarrier modulation embedding process, section 4 describes the extraction performance which analyzed the data extraction quality after being attacked and the capacity of embedded data, while the conclusion is presented in section 5.

2. Watermarking Model

The total watermark payloads of an audio watermarking system (AWS) consists of N bits hidden data as shown in figure 1. s(i) is binary watermark which has N bits content, where i is binary-based discrete time unit, since s(i) is in binary form. It is also modulated by multicarrier frequencies before embedding into the host audio. The watermark duration is 1 second with variation length of bits (N bits) to be compared and analyzed.

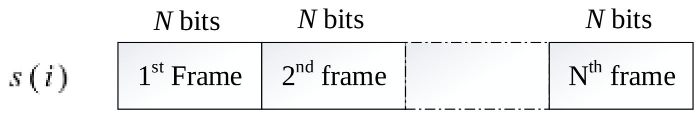

Figure 1. Watermark Payload

In the AWS, a frame is a basic unit of embedding (or extracting) a watermark bit. As shown in figure 1, after the NRZ conversion, watermark bit is modulated by multicarrier modulation and filtered by psychoacoustic filter which will be described at section 3. The filtered signal is then copied to two branch. One signal is controlled by A_1 gain. The other signal is filtered by high pass filter before controlled by A_2 gain. Finally, the watermarked audio is produced by adding those two signals after controlled by A_1 and A_2 gains with host audio.

The Audio Watermarking System (AWS) algorithm in this paper is described as following :

1. Generating the information data
2. Convert the data from binary form to NRZ form
3. Processing the binary NRZ data by multicarrier modulation
4. The modulated signal which consist of data is filtered by psychoacoustic filter to decrease the signal level to the non human auditory level.
5. The signal is multiplied by gain A_1 as primary multiplier as a controller to hidden data before embedded into the host audio.
6. The signal is also filtered by HPF and multiplied by secondary multiplier A_2 and add into the signal from step 5.
7. Embedding the signal from two branchs into the host audio by additional processing.

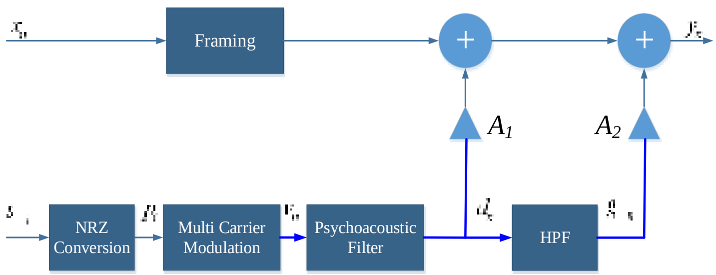

Figure 2. Processing Stages Inside Embedding Process of AWS

A communication system with multicarrier modulation transmit N_c binary values source, in parallel on N_r subcarriers with N_c frequencies of each subcarrier, f_{ij} and are converted to multicarrier symbol duration T_s. Our proposed method of audio watermarking embedding is displayed in figure 2.

The result of embedding process is given by :

$$y(n) = x(n) + A_1 d(n) + A_2 g(n) \quad (1)$$

Where
- $x(n)$ = host audio in frame-based
- A_1 = gain of $d(n)$
- A_2 = gain of $g(n)$
- $y(n)$ = watermarked audio
- $d(n)$ = the multicarrier modulated watermark data
- $d(n)$ is also the result of psychoacoustic filter process as follow :

$$d(n) = v(n) * h_P(n) \quad (2)$$

- $v(n)$ = the multicarrier modulated signal
- $h_P(n)$ = psychoacoustic filter coefficient which will be described in section 3
- $g(n)$ = the highpass signal filtered signal

$g(n)$ is the result of HPF as follow :

$$g(n) = d(n) * h_{HPF}(n) \quad (3)$$

Figure 3. Processing Stages Inside AWS Extraction

The proposed extraction process is displayed in figure 3. A frame is selected from a watermarked audio signal and then it is filtered by HPF before decoded by multicarrier demodulation. High pass filtering is used for removing low frequency signal in which most of watermark information is damaged. After demodulation, the signal is converted back to binary information by RZ conversion.

3. Multicarrier Modulation Process

Multicarrier modulation process consists of several subsystems such as : serial to parallel, copier, pulse shaper, oscillator and multiplier, and adder. The sequence of information bits is demultiplexed into several rakes. In each rake every bit is modulated by more than 1 frequency. As displayed in figure 5, the output of each rake after multicarrier modulation, as example output for i^{th} rake ($q_i(n)$) is [8]:

$$q_i(n) = p_i \sum_{j=1}^{N_c} \cos(2\pi f_{ij} n) \quad (4)$$

Thus :

$$r(n) = \sum_{i=1}^{Nr} p_i \sum_{j=1}^{N_c} \cos(2\pi f_{ij} n) \quad (5)$$

Where
- p_i = i-th binary data
- $q_i(n)$ = multicarrier modulation output of p_i
- $r(n)$ = output of multicarrier modulation for several binary data p_i
- f_{ij} = frequency used for modulation, for i-th rows and j-th column of frequency matrix as

displayed in figure 4
N_c = number of same binary data multicarrier modulated (copier)
N_r = number of rake demultiplexing (S/P) the binary data p_i

Modulation process consists of pulse shaper, multiplier, oscillator, and adder. Modulation block diagram is shown in figure 5. Frequency allocation for modulation process depends on the number of watermark data. The frequency allocation for 10, 20 and 40 bits watermark is shown in Figure 4.

Copier

S/P

10	510	1010	1510	2010	2510	....	21010	21510
60	560	1060	1560	2060	2560	....	21060	21560
110	610	1110	1610	2110	2610	....	21110	21610
160	660	1160	1660	2160	2660	....	21160	21660
210	710	1210	1710	2210	2710	....	21210	21710
260	760	1260	1760	2260	2760	....	21260	21760
310	810	1310	1810	2310	2810	....	21310	21810
360	860	1360	1860	2360	2860	....	21360	21860
410	910	1410	1910	2410	2910	....	21410	21910
460	960	1460	1960	2460	2960	....	21460	21960

Copier

S/P

10	2610	5210	....	18210
60	2660	5260	....	18260
110	2710	5310	....	18310
160	2760	5360	....	18360
210	2810	5410	....	18410
260	2860	5460	....	18460
⋮	⋮	⋮	....	⋮
2510	5110	7710	....	20710
2560	5160	7760	....	20760

Copier

S/P

10	2010	4010	6010	8010	10010	12010	14010	16010	18010	20010
60	2060	4060	6060	8060	10060	12060	14060	16060	18060	20060
110	2110	4110	6110	8110	10110	12110	14110	16110	18110	20110
160	2160	4160	6160	8160	10160	12160	14160	16160	18160	20160
210	2210	4210	6210	8210	10210	12210	14210	16210	18210	20210
260	2260	4260	6260	8260	10260	12260	14260	16260	18260	20260
310	2310	4310	6310	8310	10310	12310	14310	16310	18310	20310
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1910	3910	5910	7910	9910	11910	13910	15910	17910	19910	21910
1960	3960	5960	7960	9960	11960	13960	15960	17960	19960	21960

Figure 4. Frequencies Allocation for 10, 20, and 40 bits respectively [7]

In figure 4 the horizontal number means the frequencies (in Hz) which are used to modulate the same bit using copier to replicate the bit. The vertical number means the frequencies which are used to modulate the different bits using serial to parallel to demultiplex the bits. As example, from the middle of figure 4, the 1st bit of information will be modulated by frequencies 10, 2610, 5210, 7810, 10410, 13010, 15610, and 18210. The 2nd bit of the information will be modulated by frequencies 60, 2660, 5260, 7860, 10460, 13060, 15660 and 18260, until the last bit (20th bit) which will be modulated by frequencies 2560, 5160, 7760, 10360, 12960, 15560, 18160, and 20760. The demodulation process of the watermarked audio needs copier, multiplier, oscillator, adder, low pass filter, integrator, bit sign rounding, and parallel to serial/multiplexer. The block diagram of demodulation process is shown in Figure 6.

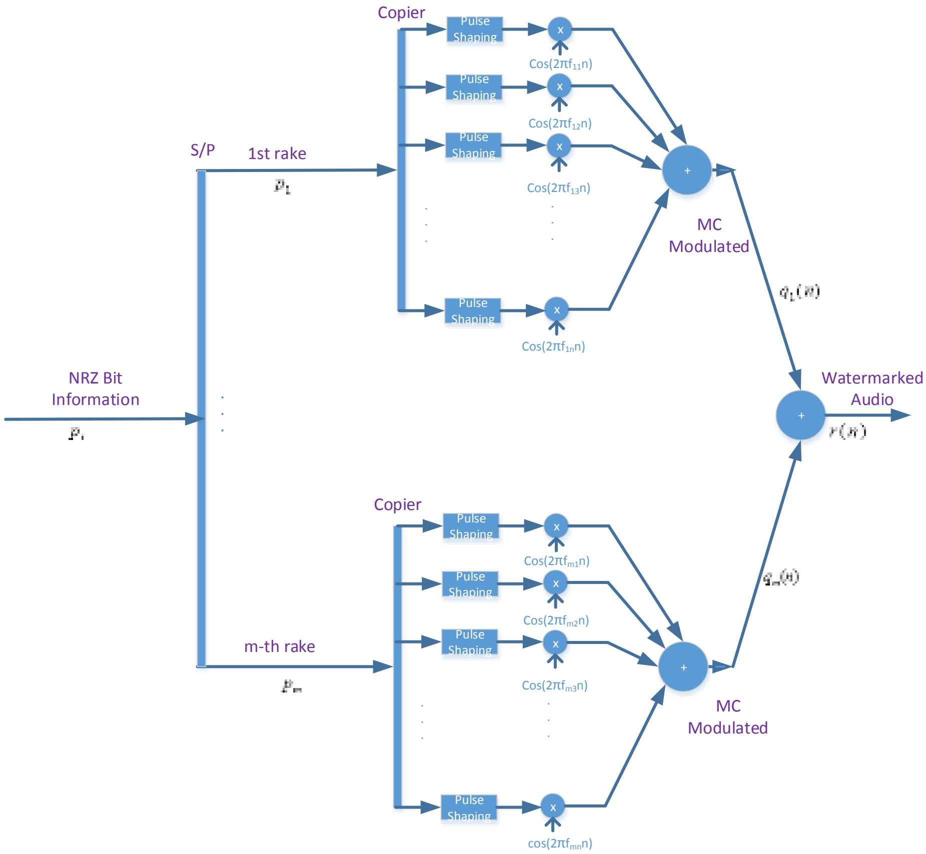

Figure 5. Multicarrier Modulation in Embedding Process of AWS

Demodulation process starts from the cosine multiplication with the same frequency as the modulation process [9] :

$$\hat{q}_i(n) = \hat{r}(n) \sum_{j=1}^{N_c} \cos(2\pi f_{ij} n) \quad (6)$$

Then it is filtered by Low Pass Filter, integrated, and detected by threshold detector [9]:

$$\hat{p}_i = \begin{cases} 1, & \sum_{i=0}^{N_s} \hat{q}_i(n) * g_{LPF}(n) \geq 0 \\ -1, & \sum_{i=0}^{N_s} \hat{q}_i(n) * g_{LPF}(n) < 0 \end{cases} \quad (7)$$

Where

N_s = Windowing/Symbol Period of demodulator

$g_{LPF}(n)$ = LPF coefficient

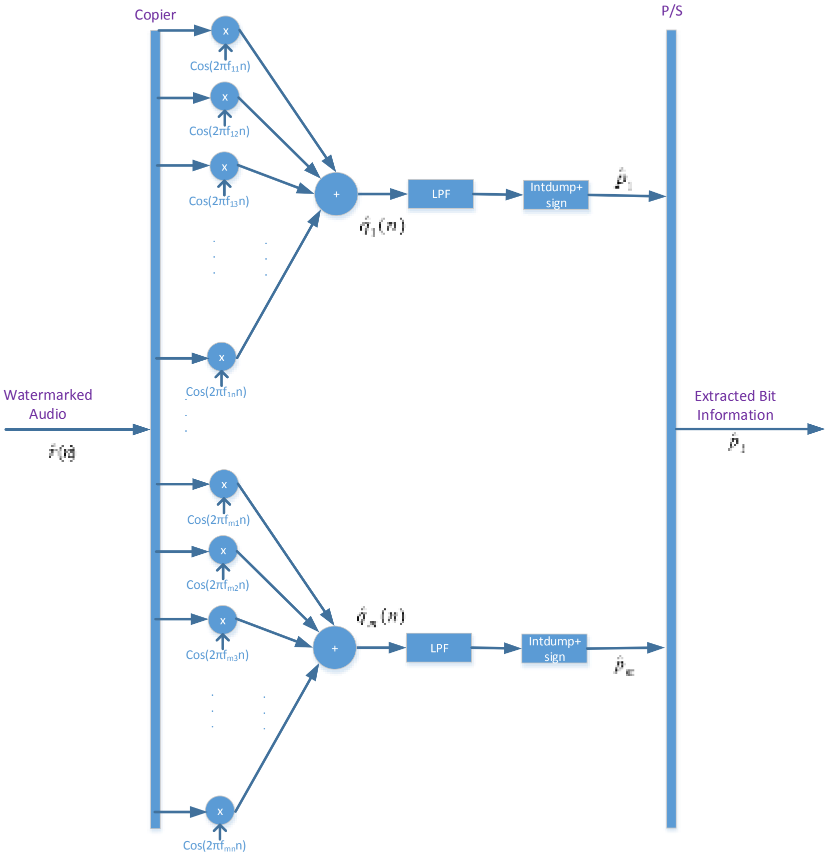

Figure 6. Multicarrier Demodulation

Psychoacoustic filtering is conducted by an Infinite Impulse Response (IIR) filter. The filter coefficients are designed by the required spectrum of psychoacoustic model and approached by pole and zero mapping design as in [10]. The psychoacoustic filter in z-domain is described by this equation :

$$H(z) = \frac{0.07-0.147z^{-1}+0.1276z^{-2}-0.053z^{-3}+0.0091z^{-4}}{1-0.15z^{-1}-0.76z^{-2}} \quad (8)$$

From this equation we can get the magnitude response which is similar with the psychoacoustic response characteristic as displayed in Figure 7. This figure describes the magnitude response of psychoacoustic filter model. In this figure the line is audibility threshold which means human can hear only and if only the signal has the amplitude bigger than the magnitude values indicated by the red line. Therefore the hidden data must have the amplitude which is less than the threshold.

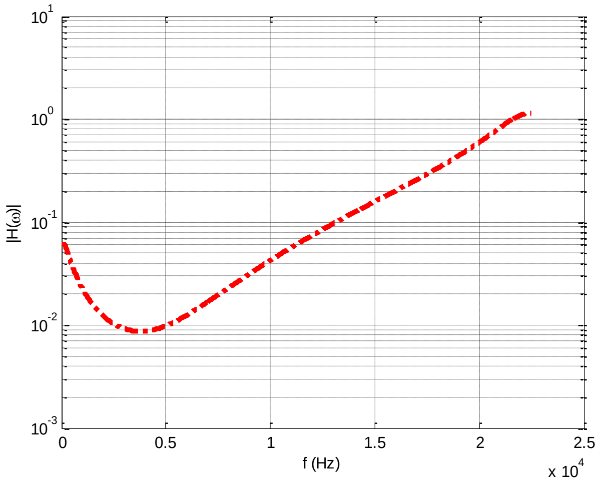

Figure 7. Magnitude response of psychoacoustic filter [10]

4. Performance Evaluation

In this section, the proposed method is evaluated in not only its imperceptibility and capacity, but also the robustness of the watermark against several audio signal processing attack. The imperceptibility of the watermark is affected by A_1 and A_2 parameter as watermark gain level in the embedding side of the audio watermarking system. The parameter representing imperceptibility are Objective Different Grade (ODG) and Subjective Different Grade (SDG). ODG and SDG will have mark as seen on table 1 as ITU-R BS.1387-1 standard about audio quality. Based on ITU-R standard, ODG is calculated via complex computation on audio signal processing based which named PEAQ (Perceptual Evaluation of Audio Quality) [11]. From table 1 we see that ODG mark has value range from -4 to 0. In other scale but on linear relation, SDG mark has value range from 1 to 5. SDG is reported by 6 respondents via listening the original and watermarked audio, then they give mark from 1 to 5 as five grade impairment scale seen in table 1. The average of their mark will be SDG per testing item.

Table 1. ITU-R five grade imperceptibility scale [11]

Subjective Scale (SDG)	Objective Scale (ODG)	Perception
1	**-4**	Very annoying
2	**-3**	Annoying
3	**-2**	Slightly annoying
4	**-1**	Perceptible but not annoying
5	**0**	Imperceptible

Aside from SDG and ODG, there is robustness parameter which has same importance with imperceptibilty parameter. The parameter is bit error rate or BER. BER is calculated by comparing the original watermark bit and the extracted or detected watermark bit. Usually, BER is calculated after extracting the watermark after the watermarked audio is attacked. But for optimizing the parameters, BER is calculated without watermarked audio attack. Parameter optimized are A_1, A_2, HPF cut off frequency in embedding (f_{co1}) and HPF cut off frequency in extraction side (f_{co2}). After the optimized parameters are selected, then the audio watermarking system with fixed parameters will be attacked by several attack for watermark robustness measurements.

A. The effect of gain level on watermarking imperceptibility

In this experiment, we will decide the value of A_1 and A_2 for acceptable audio watermarking imperceptibility. Watermark bit number used are 40 bps. HPF cut off frequency in embedding side is 10 kHz. HPF cut off frequency in extracting side is 4 kHz. The duration of the host audio is 3 s. A_1 is changed gradually from 0.00008 to 0.03, and A_2 is also set gradually from 0 to 0.01. Host audio used is "dialogue.wav" with full background music and voice during 3 s.

As shown at table 2, the imperceptibility of watermarked audio or ODG and SDG tend to increase when A_1 is decrease. A_2 also will affect the imperceptibility, when A_2 is increased, then ODG and SDG tend to decrease. But for ODG, the value is slightly fluctuative. Thus, we can choose SDG for more valid parameter for decision. Highest A_1 and A_2 for accepted imperceptibility is obtained when A_1 = 0.009 and A_2=0 (SDG=4.5), or A_1=0.01 and A_2=0.002 (SDG=3.67). We select highest combination A_1 and A_2 for next testing in order to make it robust to the watermarked audio attack, but at the same time its imperceptibility is still acceptable. For next experiment parameters used for the attack testing are A_1=0.01 and A_2=0.001. Relatively the SDG will still be acceptable for that value of A_1 and A_2. From table 2 it can be seen that there is a minimum value of A_1 for keeping the watermark extraction without error. A_1=0.0004 is minimum limit for keeping BER=0 with no attack in this experiment.

Table 2. Imperceptibility and initialization of robustness testing at host audio "dialogue.wav"

A_1	A_2	ODG	SNR	SDG	BER
0.00008	0	-0.26	49.91	5	0.15
0.00009	0	-0.3	48.89	5	0.16
0.0001	0	-0.28	47.98	5	0.12
0.00011	0	-0.3	47.15	5	0.12
0.0002	0	-0.44	41.95	5	0.08
0.0003	0	-0.6	38.44	5	0.08
0.0004	0	-0.65	35.94	5	0
:	:	:		:	:
:	:	:		:	:
0.008	0	-2.24	10.34	4.50	0
0.009	0	-2.3	9.42	4.50	0
0.01	0	-2.2	8.62	3.50	0
0.02	0	-2.2	4.1	2.83	0
0.03	0	-2.22	2.3	2.50	0
0.01	0.01	-2.79	4.1	2.83	0
0.01	0.008	-2.2	4.68	2.83	0
0.01	0.006	-2.52	5.38	3.33	0
0.01	0.004	-2.2	6.24	3.33	0
0.01	0.002	-2.26	7.3	3.67	0

B. The effect of watermark payload on watermarking performance

In this experiment, audio file used for embedding and extraction is "dialogue.wav". Watermark data is generated on randomly uniform distribution. Watermark bit number is set to be 10 bps and 40 bps for performance comparison. A_2 is set to be 0. Only A_1 is changed from 0.00006 to 0.0004. The result is displayed on table 3. It can be seen that limit of perfect BER and error BER will increase since the watermark payload increase. When bit number is 10 bps, BER will be no longer zero at A_1=0.00008, but when bit number is 40 bps, BER will be no longer zero at A_1=0.0004. This means that the robustness of watermark will decrease since the watermark payload increases. The payload is inversely proportional with the robustness.

Table 3. Bit number effect on imperceptibility and robustness at "dialogue.wav"

Bit Number	A1	ODG	SDG	BER
10 bps	0.00006	-0.21	5	0.1
	0.00007	-0.25	5	0.07
	0.00008	-0.26	5	0.03
	0.00009	-0.27	5	0
	0.0001	-0.29	5	0
	0.00011	-0.31	5	0
	0.0002	-0.44	5	0
	0.0003	-0.59	5	0
	0.0004	-0.89	5	0
40 bps	0.00006	-0.2	5	0.19
	0.00007	-0.26	5	0.09
	0.00008	-0.26	5	0.15
	0.00009	-0.3	5	0.16
	0.0001	-0.28	5	0.12
	0.00011	-0.3	5	0.12
	0.0002	-0.44	5	0.08
	0.0003	-0.6	5	0.08
	0.0004	-0.65	5	0

C. *Testbed Result Performance*

The testbed procedure by several attack types has several points of testing which generally consists of lowpass filtering, bandpass filtering, noise addition, resampling, time scale modification, linear speed change, pitch shifting, multi band equalizing, echo addition, and MP3 compression. The attack type and their description are displayed in table 4. Cut off frequency for LPF attack are set to 6, 9, 12, 16 kHz. High cut off frequency for BPF attack are also set to 6, 9, 12, 16 kHz. Noise additive will add the noise to the watermarked audio in cascade additional between white noise and pink noise. The level noise is set to 20 dB below average watermarked audio power. Resampling attack is set to 22.05 kHz, 16 kHz, and 11.025 kHz. Linear speed change attack are set to -15%, -10%, -5%, 5%, 10%, and 15%. Multi band equalizer attack will have 10 band graphic and consists of band frequency (Hz) : [31 62 125 250 500 1 kHz 2 kHz 4 kHz 8 kHz 16 kHz] and gain (dB) [-6 +6 -6 +6 -6 +6 -6 +6 -6 +6]. Echo addition attack will have maximum delay 100 ms and feedback coefficient around 0.3. And last attack is MP3 compression which has rate : 32 kbps, 64 kbps, 128 kbps, 192 kbps, and 256 kbps. The host audio used for embedding, attacking, and extraction consist of 5 audio files, that is "dialogue.wav", "fleetwd.wav", "Moonriver_Mancini.wav", "mouth_harmonica.wav", and "Sax_Piano.wav". Each host audio file has duration 3 s. Watermark data is generated on randomly uniform distribution at 10 and 40 bps payload.

Table 4. Testbed list as watermarked audio attack

Item	Type of attack	Attack Description
1	Low pass filter	cut off frequency = 9, 12, 16 kHz, second order butterworth filter
2	Band pass filter	cut off frequency = 100 Hz - 9 kHz, 100 Hz - 12 kHz, 100 Hz - 16 kHz, 2nd order Butterworth filter
3	Noise addition	Adding white and pink noise with constant level of 20 dB lower than total averaged music power
4	Changing the sample rate	44.1 kHz -> 22.05 kHz, 44.1 kHz -> 16 kHz, 44.1 kHz -> 11.025 kHz
5	Linear speed change	-15%,-10%,-5%, 5%, 10%,15%
6	Multi-band equalization	10-band graphic equalizer with the characteristics listed below: Freq.[Hz]: 31 62 125 250 500 1kHz 2kHz 4kHz 8kHz 16kHz Gain[dB]: -6 +6 -6 +6 -6 +6 -6 +6 -6 +6
7	Echo addition	Maximum delay: 100 ms Feedback coefficient: around 0.3
8	MP3 Compression	32, 64, 128, 192, 256 kbps

The used parameters for testbed : A_1=0.01, A_2=0.001, HPF cut off frequency in embedding side is 10 kHz in order to keep good imperceptibility, HPF cut off frequency in extracting side is 4 kHz to keep the good robustness, especially when the watermarked audio faces low pass filtering attack with 6 kHz cut off frequency. Watermark payload used are 10 bps and 40 bps. The overall results displayed in table 5 are average of bit error rate from each file in every attack type and every parameter of attack.

Table 5. Testbed result

Item	Type of attack	Payload (bps)	Parameters	Average BER
1	Low pass filter	10	fco= 6, 9, 12, 16 kHz	0.34, 0.27, 0.09, 0.007
		40	fco= 6, 9, 12, 16 kHz	0.47, 0.37, 0.17, 0.03
2	Band pass filter	10	fco=100 Hz – 6/9/12/16 kHz	0.34, 0.27, 0.09, 0.007
		40	fco=100 Hz – 6/9/12/16 kHz	0.48, 0.37, 0.17, 0.03
3	Noise addition	10, 40	White and pink noise with 20 dB lower than averaged audio power	0, 0
4	Changing the sample rate	10	22.05, 16, 11.025 kHz	0.0067, 0.0067, 0
		40	22.05, 16, 11.025 kHz	0.005, 0, 0.002
5	Linear speed change	10	-15, -10, -5, 5, 10, 15 %	0, 0, 0, 0, 0, 0
		40	-15, -10, -5, 5, 10, 15 %	0, 0, 0, 0, 0, 0
6	Multi-band equalization	10, 40	10-band graphic equalizer with the characteristics listed below: Freq.[Hz]: 31 62 125 250 500 1kHz 2kHz 4kHz 8kHz 16kHz Gain[dB]: -6 +6 -6 +6 -6 +6 -6 +6 -6 +6	0, 0
7	Echo addition	10, 40	Maximum delay: 100ms Feedback coefficient: around 0.3	0, 0
8	MP3 Compression	10	32, 64, 128, 192, 256 kbps	0.47, 0.09, 0, 0, 0
		40	32, 64, 128, 192, 256 kbps	0.48, 0.2, 0, 0, 0

From table 5, it can be seen that overall audio watermarking in multicarrier-based with the optimized parameter as explain in subsection IV.A obtains good result. The perfect robustness results with BER 0 are obtained in noise addition attack, linear speed change attack, multi band equalization attack, and echo addition attack. Audio watermarking robustness in LPF and BPF attack is similar. The robustness tends to be better when the cut off frequency of LPF and BPF increases. The watermark robustness from last attack, MP3 compression, has acceptable result, except at compression rate 32 kbps. The robustness from MP3 compression attack is perfect when MP3 compression rate is more closely with 128 kbps. Especially for 10 bps payload, this technique of audio watermarking has acceptable robustness with MP3 compression attack for MP3 compression rate more closely with 64 kbps.

Table 6. Robustness comparison with different method

Scheme	**LPF**	**Resampling**		**MP3 Compression**	
		11.025 kHz	**22.05 kHz**	**64 kbps**	**128 kbps**
[12]	0% (18 kHz)	NA	NA	9%	10%
[13]	1.43% (6 kHz)	NA	0	1.43%	NA
[14]	NA	NA	0	NA	3.30%
[15]	NA	NA	0	NA	2.93%
Proposed (40 bps payload)	3% (16 kHz)	0.20%	0.50%	20%	0%
Proposed (10 bps payload)	0.7% (16 kHz)	0.67%	0%	9%	0%

Table 6 displays robustness comparison with last different method of frequency domain based audio watermarking, but not all method described the robustness of the same attack, thus several attacks obtained NA (not available) robustness. Mehdi in [12] used Fibonacci as sequence for embedding watermark into host audio in frequency domain by FFT. His method obtained perfect watermark with zero BER when the watermarked audio was attacked by LPF, but the cut off frequency was set to 18 kHz. Mehdi obtained 10% BER when the audio watermarked was attacked by 128 kbps MP3 compression. Yiqing [13] used FFT as transform domain and psychoacoustic model with gammatone filter for embedding the watermark. She also used artificial intelligence after extracting the watermark to get binary based watermark data, thus she obtained the robustness better than ours in LPF attack, and MP3 compression. Pranab [14] used DWT-DCT-SVD method to embed and extract the watermark in audio watermarking. And he also published another method in [15] using FFT-SVD-CPT method. Anyway, our method has reached perfect robustness when the watermarked audio files are attacked by 128 kbps MP3 compression, while the other method couldn’t reach that performance.

5. Conclusion

The testing result of multicarrier modulation shows that information bit hidden in the host audio could reach up to 40 bps and the imperceptibility level is still acceptable, due to MOS result is more than 4. The testbed procedure result gives the perfect robustness in noise addition attack, linear speed change attack, multi band equalization attack, and echo addition attack. And the robustness of audio watermarking is still acceptable with MP3 compression attack and filtering for cut off frequency up to 16 kHz, due to BER is lower than 10%. Comparing with other current method, this proposed method has perfect robustness or zero BER when the attack type used 128 kbps MP3 compression.

6. Acknowledgements

This research was partially supported by ITB WCU grant 2016 and RISET ITB 2016

7. Bibliography

[1]. R. Van Nee and R. Prasad, OFDM for Wireless Multimedia Communications. London: Artech House, 2000

[2]. N. Marchetti, M. I. Rahman, and S. Kumar, New Directions in Wireless Communications Research. 2009

[3]. S. Maiti, A. Roy, T. S. Das, and S. K. Sarkar, "OFDM Based High Capacity Information Hiding in Grey Scale Image," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 1, no. 6, pp. 192–195, 2012

[4]. R. Amirtharajan, K. Thenmozhi, and R. J. B. Balaguru, "Multi Carrier Steg against Omni Attacks," *International Journal of Computer Applications*, vol. 5, no. 11, pp. 35–40, 2010

[5]. A. V Shishkin, "OFDM-based Audio Watermarking for Electronic Radiotelephone Identification," *Design & Test Symposium (EWDTS), 2010 East-West*, pp. 190–194, 2010

[6]. J. J. Garcia-Hernandez, R. Parra-Michel, C. Feregrino-Uribe, and R. Cumplido, "High payload data-hiding in audio signals based on a modified OFDM approach," *Expert Systems with Applications*, vol. 40, no. 8, pp. 3055–3064, 2013

[7]. G. Budiman, A. B. Suksmono, and D. H. Shin, "A Multicarrier Modulation Audio Watermarking System," in *International Conference on Electrical Engineering and Informatics* 2015, 2015, no. 2, pp. 164–171

[8]. K. Fazel and S. Kaiser, Multi-Carrier and Spread Spectrum Systems. 2003

[9]. R. E. Z. (University of Colorado) and W. H. T. (Virginia P. I. and S. University), Principles of Communications : *System, Modulation and Noise, 6th ed. John Wiley and Sons*, 2010

[10]. D. J. M. Robinson, "the Human Auditory System," 107th convention of the Audio Engineering Society, pp. 1–13, 1999

[11]. R. ITU-R, RECOMMENDATION ITU-R BS.1387-1 Method for objective measurements of perceived audio quality, vol. BS.1387-1. 1998

[12]. M. Fallahpour and D. Megías, "Audio Watermarking Based on Fibonacci Numbers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 8, pp. 1273–1282, 2015

[13]. Y. Lin and W. H. Abdulla, Audio Watermark, A Comprehensive Foundation Using Matlab. Auckland: Springer, 2015

[14]. P. K. Dhar and T. Shimamura, Advances In Audio Watermarking Based on SVD. New York, London: Springer, 2015

[15]. P. K. Dhar and T. Shimamura, "Audio watermarking in transform domain based on singular value decomposition and Cartesian-polar transformation," *International Journal of Speech Technology*, vol. 17, no. 2, pp. 133–144, 2014

Gelar Budiman received B.S. and M.S. degree in electrical engineering from Sekolah Tinggi Teknologi Telkom (STTT), Bandung, Indonesia. He has been working as a lecturer in Telkom University since 2008. He has been currently taking doctoral degree in School of Electrical Engineering and Informatics, Bandung Technology Institute (ITB), since 2015. His research interests are in the areas of wireless communication, signal processing, and watermarking.

Andriyan Bayu Suksmono (M'02–SM'08) received the B.S. degree in physics and the M.S. degree in electrical engineering from the Bandung Institute of Technology (ITB), Indonesia, and the Ph.D. degree in engineering from the University of Tokyo, Japan, in 1990, 1996 and 2002, respectively. He joined ITB as an Instructor (1996–2005), Associate Professor (2005–2009), and Professor (2009-present) at the School of Electrical Engineering and Informatics, ITB. His main research interests are signal processing and imaging.

Donny Danudirdjo received the B.S. and M.S. degrees in electrical engineering from the Bandung Institute of Technology (ITB), Bandung, Indonesia, in 2005 and 2007 and the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 2013.

Since 2014, he has been with the School of Electrical Engineering and Informatics, ITB. His current research interests include interferometric synthetic aperture radar image analysis and fractal-related topics in signal and image processing.

Koredianto Usman was born in Sumatera Selatan Indonesia October 2, 1975. His obtained his bachelor from Electrical Engineering Bandung Institute of Technology, Indonesia, 1999, and master degree from Munich Institute of Technology, Germany in 2001. He is now a researcher at School of Electrical Engineering and Informatics, Bandung Technology Institute (ITB).

Dong Hwan Shin received the MS degree and Ph.D. in electronics engineering from University of Seoul, Korea. From 1992 to 1994, he was a member of the LG Electronics Inc. He worked as a senior researcher in the Korea Sports Science Institute from 1996 to 2000. He has been currently a chief manager of MarkAny Inc. from 2000. Her research interests are in the areas of copyright protection, fingerprinting, watermarking and machine learning.