

PAPER • OPEN ACCESS

## Multichannel audio steganography based on MPEG surround using direct sequence spread spectrum

To cite this article: M Tomas *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **602** 012011

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

# Multichannel audio steganography based on MPEG surround using direct sequence spread spectrum

M Tomas<sup>1</sup>, Baharuddin<sup>1</sup>, and I Elfriti<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Andalas University, Padang, Indonesia

E-mail: ikhwana@eng.unand.ac.id

**Abstract.** Audio steganography is a technique for embedding hidden message on the audio signal. Several techniques are currently available, proposed as methods to hide secret messages on audio signals and integrated into the audio encoding system. In this paper, a data hiding technique is proposed working based on MPEG Surround (MPS), a multichannel audio encoding standard that is very popular for spatial or three-dimensional (3D) audio coding. Direct sequence spread spectrum (DSSS) is integrated with MPS to embed a secret message into the downmix signal of multichannel audio that generated from MPS encoder. The result of experiments shows that the audio that produced by the proposed system is still in acceptable quality and signal synchronization can run smoothly. The mean value of the signal to noise ratio (SNR) of the audio signal is 16.67 dB. The secret message can be successfully extracted with a mean value of bit error rate (BER) of 3.38%, and normalized correlation (NC) is 95%.

## 1. Introduction

The rapid development of internet network and computer technology today is making everyone can easily access, copy and modify digital multimedia products, especially digital audio. So that it makes the distribution of digital music content more difficult to control. This causes a lot of privacy issues and violations of copyright. Therefore, digital data encryption is very important [1]. A technique that can be used to secure digital multimedia audio is steganography [2].

Many researchers have been carried out which related to how to embed secret messages into audio signals using the steganography method [3]. They are Least Significant Bit (LSB) method [4,5], Parity Coding method [6], Phase Coding method [7], Echo Data Hiding method [8], and Direct Sequence Spread Spectrum method (DSSS) [9]. The most preferred method of steganography is Direct Sequence Spread Spectrum because this system is the most effective and robust steganography method in storing confidential data in digital audio [10]. The medium where information is embedded is called an audio cover, which can be used MPEG-1 Layer 3, Advance Audio Codec (AA and original audio (wave) [11]. However, the research that has been done on audio steganography is still focused on mono channel audio and stereo channel.

The research using multichannel audio technology is still limited. This technology aims to produce audio output that its ability to represent the sound produced around the listener based on human perception of the audio itself so that the result of audio output gives a real impression to the audience [12,13]. The research development on multichannel audio today is used MPEG Surround technology (MPS). However, its still developing on improving the quality of the resulting multichannel audio [14–20]. Therefore the author is motivated to research by designing and implementing multichannel audio

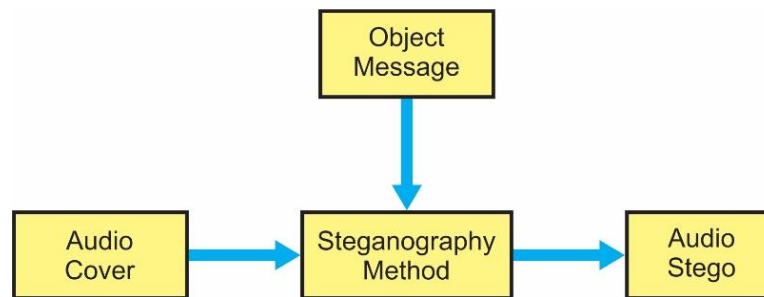


steganography on MPEG Surround (MPS) using the Direct Sequence Spread Spectrum (DSSS) method. MPS will be integrated with DSSS to embed a secret message into multichannel audio cover. Furthermore, from the test of results objectively analyzed the value of Signal-to-Noise Ratio (SNR), Bit Error Rate (BER) and Normalized Correlation (NC).

## 2. Brief Overview of Audio Steganography and MPEG Surround

### 2.1. Audio Steganography

Steganography is a technique and the art of concealing information on another medium to keep the information hidden [21]. The steganography application can be used to embed object messages into audio cover. So that it needs an algorithm that can modify digital objects into new objects. This algorithm is called an embedding algorithm. While the algorithm used to retrieve messages that have been embedded in the medium (audio stego signal) is called the retrieving or extracting algorithm. Both algorithms are represented of the system that contained steganography method called embedding encoder and extracting decoder. Audio steganography can be used in the time domain, frequency domain, and compression domain. This work uses a technique in the time domain.



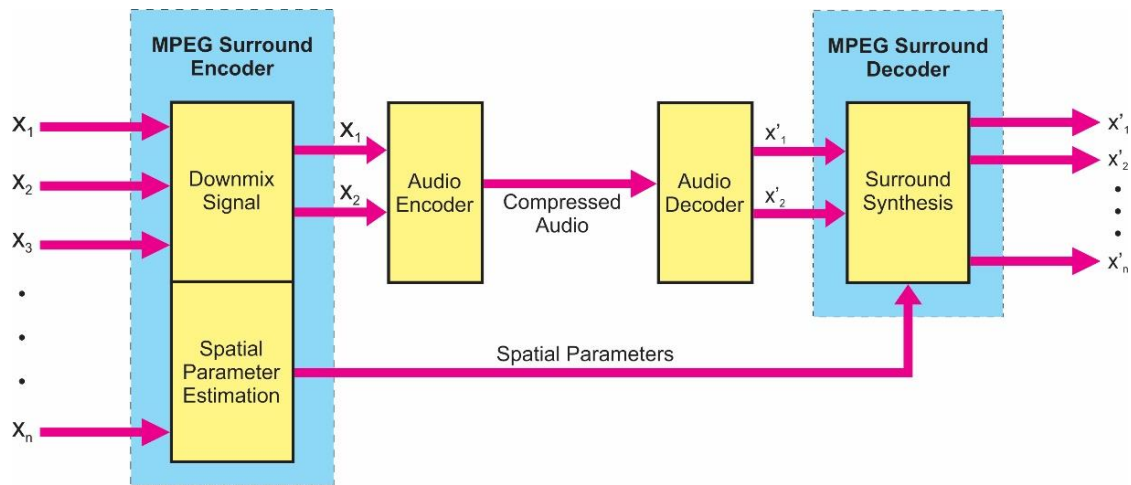
**Figure 1.** A basic block diagram of audio steganography

### 2.2. Direct Sequence Spread Spectrum Method

There is a method of audio steganography that is the most effective and robust in storing confidential data in digital audio called Direct Sequence Spread Spectrum (DSSS) [10]. DSSS is a method of embedding data by spreading confidential data along the audio cover signal. Spread spectrum has very high advantages important, its resistance to jamming and interference. If the signal that is made is damaged in the middle of the road, the information that is conveyed can still be perceived. Object message in text file form is converted into binary form. For spreading data of object message, DSSS generate PN sequence key and then multiplied with text binary form which the result is called spread spectrum data.

### 2.3. MPEG Surround

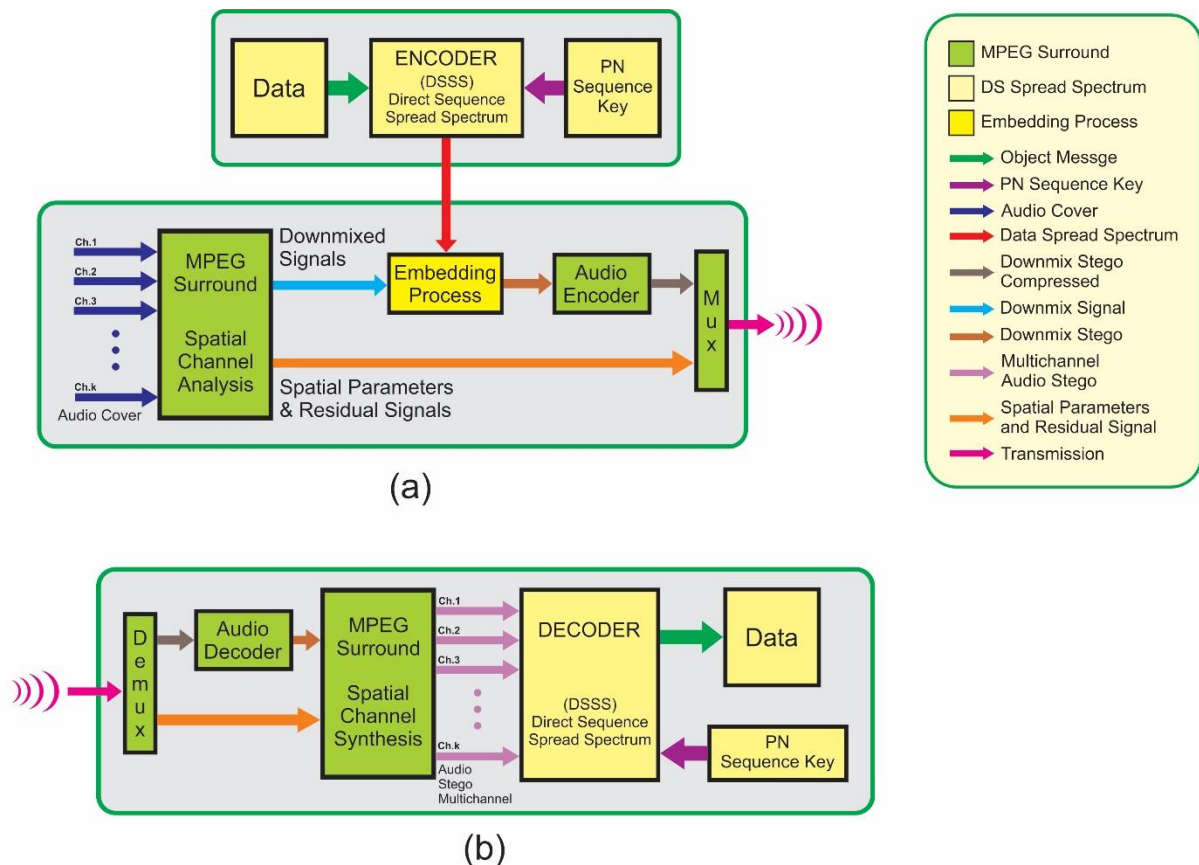
MPEG Surround (MPS) is one of the spatial audio codings that has been standardized by the International Organization for Standardization (ISO). MPEG Surround coding exploits our ability to perceive sound in three dimensions and captures that perception in a compact set of parameters. It exploits inter-channel differences in level, phase, and coherence equivalent to the Interaural Level Differences (ILD), Interaural Time Difference (ITD) and Interaural Coherence (IC) cues to capture the spatial image of a multi-channel audio signal relative to a transmitted downmix signal and encodes these cues in a very compact form such that the cues and the transmitted signal can be decoded to synthesize a high-quality multi-channel representation. The multichannel audio as the input on MPEG Surround encoder will be analyzed by hybrid filterbank and then decomposed to be a downmix signal, residual signals, and spatial parameters. The downmix signal is typically stereo but could also be mono that is derived from the multichannel audio input signal. In this work, the downmix signal will be written to be a wave audio signal and then becomes an audio cover in steganography that it will be embedded a secret message in text form. This is illustrated in Figure 2.



**Figure 2.** A basic block diagram of MPEG Surround

### 3. Audio Steganography on MPEG Surround

#### 3.1. System Architecture



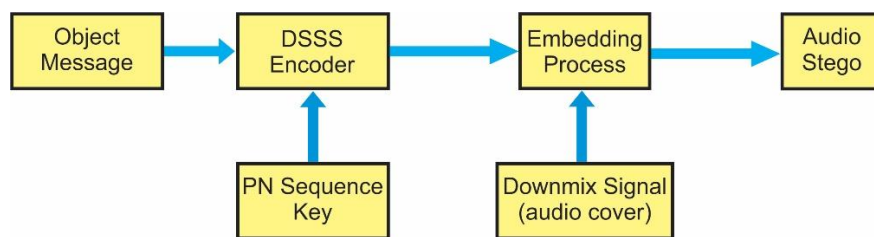
**Figure 3.** (a) Embedding process of steganography (b) Extracting process of steganography

Based on figure 3, the system architecture is designed by the integration between DSSS and MPS, which is called the DSSS-MPS multichannel audio steganography. Multichannel audio becomes audio input on MPEG Surround encoder. Then the downmix signal, residual signal, and spatial parameters will be extracted from the MPS encoder. The secret messages as object messages are multiplied by PN

Sequence key with X-OR operations on the DSSS encoder so that the result will generate spread spectrum data that will be embedded into audio cover signal. The downmix signal as an audio cover is embedded with object message in embedding process which will produce an audio signal that has an embedded message that is called downmix stego signal. It is encoded by Advance Audio Codec (AAC) encoder with a total of bit rate more than 200 kb/s. The output signal of downmix stego from AAC encoder along with residual signal and spatial parameters are forwarded to the multiplexer and then transmitted into the MPEG Surround decoder as the receiver. MPS decoder is converting back audio stego signal into a waveform by AAC decoder. The residual signal and spatial parameters will be synthesized by hybrid filterbank of MPS decoder for generating a multichannel audio signal and then forwarded into the DSSS Decoder for extracting the object message by multiplied with the PN sequence key.

### Embedding Process

The embedding process is to embed the text file into the downmix signal as audio cover in the audio waveform. In this work, object message is converted into binary form firstly. The object messages in binary form will be embedded into along of downmix signal as audio cover. In this stage more details there are three processes, the PN sequence key generation process, the encoding process, and affixing process. This is illustrated in Figure 4 below.



**Figure 4.** A basic block diagram of the embedding process for downmix signal as audio cover

#### 3.1.1. PN Sequence Key Generation

The frequency of Pseudo Noise Sequence key or commonly called chip rate affects payload cover-object. The stored audio data has a sampling frequency of 48 kHz. Consequently, the PN Sequence frequency should be the same or more than 48 kHz. If we want to send message information with 32-bit length, it means we send 48000 sampling information for each bit. In this paper, the audio cover that is used as a sampling frequency of 48 kHz with a duration of 12 seconds. So that the number of audio cover samples is 576 kHz. If the amount of length of information binary is  $L_{bit}$  and the number of audio cover samples is  $L_{sig}$ , then we can get the number of binary segments in samples is  $N$  with the following equation.

$$N = \frac{L_{sig}}{L_{bit}} \quad (1)$$

$$S_{bit} = \frac{L_{sig}}{N} \quad (2)$$

$S_{bit}$  is the number of binary samples in one segment frame. So that for embedding all binary of object message into the audio cover, the PN sequence key is generated as many as the number of sample of audio cover signal. Then PN sequence is multiplied to spread spectrum data with X-OR operation.

### 3.1.2. Encoding Process

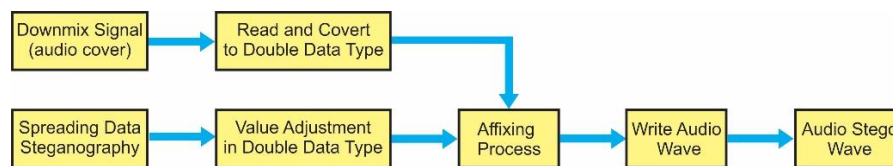
The encoding process is carried out to produce spread spectrum data which will be embedded into an audio cover based on the DSSS method. Digital text files as object messages are first converted into binary form. Then the object message in binary form is encoded by using PN Sequence key that has been generated.

### 3.1.3. Affixing Process

Downmix audio signal is used as an audio cover in the waveform. Before the spread spectrum data is added to audio cover, the entire main data of wave read firstly and then divided into several frame segments. Every single frame has an audio sample value. The spread spectrum data of object message is affixed to the sample according to the value audio cover scale. Each frame contains audio sample data that has a value scale audio between +1 to -1. The spread spectrum data has an audio value adjustment which aims to convert its value into audio value. At the output of the affixing process, the spread spectrum data is added to the audio cover resulting in:

$$y'(t) = y_i(t) + \alpha.w_i \quad (3)$$

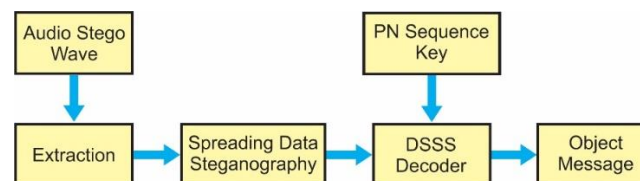
Where  $\alpha$  is the steganography coefficient that will be inserted into the audio cover signal,  $y_i'(t)$  is an audio cover, and  $w_i(t)$  is spread spectrum data in time domain. The result of the affixing process is called audio stego  $y'(t)$  which is converted into the waveform. This is illustrated in Figure 5.



**Figure 5.** Block diagram scheme of affixing process

### 3.2. Extraction Process

The extraction process of spread spectrum data is done to separate it from digital audio wave cover files and rearrange it becomes a text file as a retrieved message. The process begins with the decomposition process and then followed by the decoding process.



**Figure 6.** Block diagram extraction process

PN sequence key that is used in the extraction stage is same with PN sequence key that is generated during the affixing process. PN sequence functions as a key for the decoding process on the spread spectrum data so that it can be reshaped become text file.

#### 3.2.1. Decomposition

The decomposition process takes the audio stego wave for dividing into frame segments as much as the length of the object message. Spread spectrum data is represented in each frame segment, and then for extracting the object message from audio stego, the value of audio stego in each frame segment is multiplied by PN sequence key in the DSSS decoder.

### 3.2.2. Decoding Process

The decoding process serves to change the spread spectrum data becomes object message. Decoding process that is applied in the DSSS method has the opposite steps of the encoding process during the affixing stage. So that, spread spectrum data can be changed back into object message data, it must be multiplied by PN Sequence key. The object message data that has been obtained is converted back into the form of a letter character so that the message content can be read.

### 3.3. Experiment to evaluate the Audio Steganography System

The most important for audio steganography is the insertion of object message data should not change the sound quality of the audio signal that is used as a cover and also the character of the retrieved message is complete. So that, the secret data that listeners and the object message can not detect embedded can be read. Parameters that are used to determine the quality of audio cover that has been embedded data are signal to noise ratio (SNR), bit error rate (BER) and normalized correlation (NC). This parameter can be calculated by the following equation.

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} [x'(n) - x(n)]^2} \right\} \quad (4)$$

$$BER(x, x') = \frac{(x \oplus x')}{N} \quad (5)$$

$$NC(x, x') = \frac{(x + x')}{\sqrt{x^2} \sqrt{x'^2}} \quad (6)$$

Where  $x$  is the original audio signal data and  $x'$  the steganography data signal and  $N$  is the length of the audio cover and  $\oplus$  is X-OR operation.

## 4. Result and Discussion

This work aims to design and implement multichannel audio steganography on MPS using DSSS in the time domain. Testing of the result is using Matlab simulation by embedding text with five kinds of variations and five samples of the audio cover signal. The audio cover signals are multichannel audio original in waveform with a sampling frequency of 48 kHz, duration for 12 seconds and five channels. Each audio cover has a size of 5.49 Mb with a bit rate of 840 kb/s and a bit depth of 16 bit. Advance Audio Codec (AAC) that is used on MPEG Surround is converting the audio cover signal from a wave into compressed form with five variations of bit rate more than 200 kb/s. Information of samples is described in Table 1 and 2.

**Table 1.** Information of Sample Audio Cover in Audio Steganography System

Audio Cover File	Duration (second)	Size (Mb)	Bit Rate (kbps)	Sample Rate (Hz)	Bit Depth (a bit)	Number of Channel
drum.wav	12	5.49	840	48000	16	5
applaus.wav	12	5.49	840	48000	16	5
talking.wav	12	5.49	840	48000	16	5
laughter.wav	12	5.49	840	48000	16	5
vivaldi.wav	12	5.49	840	48000	16	5



**Table 2.** Information of Sample Object Message in Audio Steganography System

Object Message	Number of Char	Length (bit)	Binary Data
ABCD	4	32	01000001 01000010 01000011 01000100
ABCDE	5	40	01000001 01000010 01000011 01000100 01000101
ABCDEF	6	48	01000001 01000010 01000011 01000100 01000101 01000110
ABCDEFG	7	56	01000001 01000010 01000011 01000100 01000101 01000110 01000111
ABCDEFGH	8	64	01000001 01000010 01000011 01000100 01000101 01000110 01000111 01001000

#### 4.1. Signal to Noise Ratio (SNR)

**Table 3.** Result in The Value of SNR of Experiments

Object Message	AAC bitrate (Kbps)	Variations of Audio Cover					Average (dB)
		drum (wav)	applause (wav)	talking (wav)	laughter (wav)	vivaldi (wav)	
ABCD	288	15.91	14.37	16.29	15.41	16.71	15.74
	320	16.17	14.91	16.79	15.83	17.15	16.17
	400	16.51	15.73	17.33	16.41	17.62	16.72
	480	16.91	16.65	17.81	17.30	18.19	17.37
	560	16.98	16.81	17.88	17.43	18.29	17.48
ABCDE	288	15.89	14.37	16.28	15.41	16.71	15.73
	320	16.16	14.91	16.77	15.83	17.13	16.16
	400	16.50	15.73	17.32	16.41	17.61	16.71
	480	16.91	16.65	17.81	17.30	18.18	17.37
	560	16.97	16.81	17.89	17.43	18.28	17.47
ABCDEF	288	15.90	14.37	16.25	15.41	16.72	15.73
	320	16.17	14.91	16.75	15.83	17.15	16.16
	400	16.51	15.73	17.29	16.41	17.62	16.71
	480	16.91	16.64	17.78	17.30	18.20	17.37
	560	16.97	16.81	17.86	17.43	18.29	17.47
ABCDEFG	288	15.88	14.37	16.28	15.41	16.71	15.73
	320	16.15	14.91	16.77	15.83	17.14	16.16
	400	16.50	15.74	17.31	16.41	17.62	16.71
	480	16.90	16.65	17.81	17.30	18.19	17.37
	560	16.96	16.82	17.87	17.43	18.29	17.47
ABCDEFGH	288	15.89	14.37	16.27	15.42	16.71	15.73
	320	16.15	14.91	16.76	15.84	17.14	16.16
	400	16.49	15.74	17.29	16.42	17.61	16.71
	480	16.89	16.65	17.79	17.31	18.18	17.36
	560	16.95	16.82	17.87	15.42	18.28	17.07



Based on table 3, the results of the experiment obtained the mean values of SNR for all variations of 16.67 dB. This SNR value shows the margin of the signal to noise ratio level with a good classification. Its mean that the signal synchronization can run smoothly. This result shows that is all of the audio covers can be a medium for embedding data text as object message. The mean value of SNR for all variations of text object messages in each variation of bitrate is constantly because in this algorithms data is being spread into whole audio cover signal. So that, the whole signal is used for embedding, no matter how short the message. Variations of the bit rate of Advance Audio Codec (AAC) are affecting the value of SNR if bitrate gets bigger, so the value of SNR also will get bigger although the size of the text is getting bigger.

#### 4.2. Bit Error Rate (BER) and Normalized Correlation (NC)

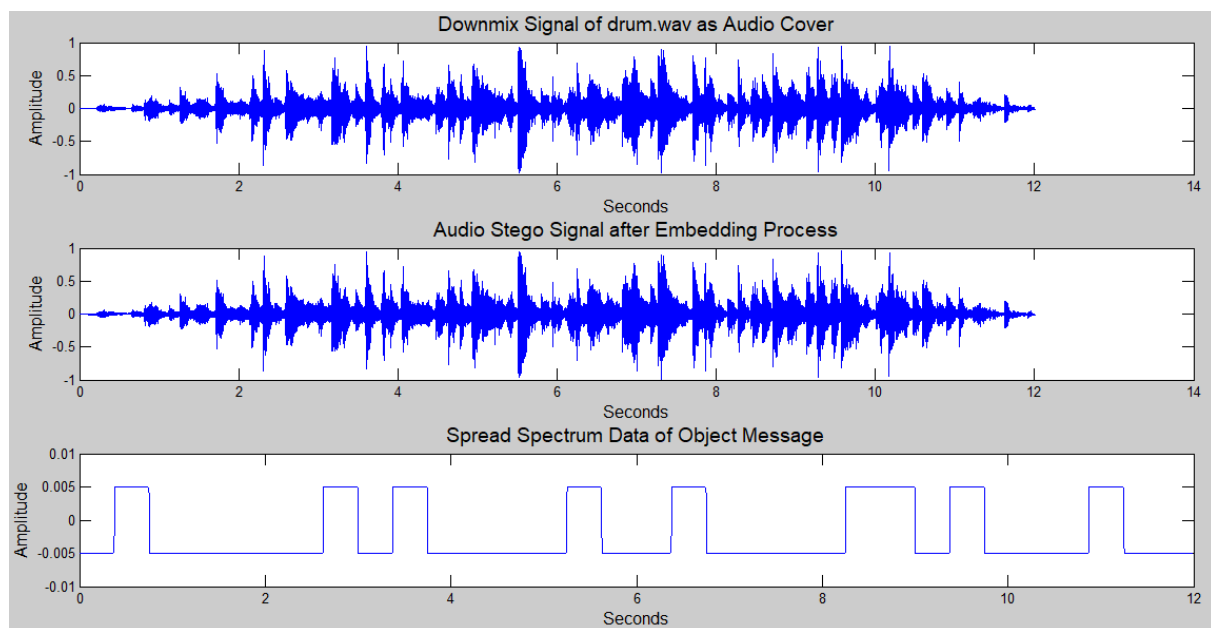
Based on Table 4 below, the results of an experiment for all variations show the mean values of bit error rate (BER) of 3.38 and normalized correlation (NC) of 0.95. This result shows that object message that is embedded in audio cover can be extracted.

**Table 4.** Result in The Value of BER and NC of Experiments

Object Message	AAC bitrate (Kbps)	Variations of Audio Cover									
		drum (wav)		applaus (wav)		talking (wav)		laughter (wav)		vivaldi (wav)	
		BER	NC	BER	NC	BER	NC	BER	NC	BER	NC
ABCD	288	0	1	0	1	0	1	0	1	0	1
	320	0	1	0	1	0	1	0	1	0	1
	400	0	1	0	1	0	1	0	1	0	1
	480	0	1	0	1	0	1	0	1	0	1
	560	0	1	0	1	0	1	0	1	0	1
ABCDE	288	0	1	0	1	0	1	0	1	0	1
	320	0	1	0	1	0	1	0	1	0	1
	400	0	1	0	1	0	1	0	1	0	1
	480	0	1	0	1	0	1	0	1	0	1
	560	0	1	0	1	0	1	0	1	0	1
ABCDEF	288	0	1	0	1	0	1	0	1	22.92	0.69
	320	0	1	0	1	0	1	0	1	22.92	0.69
	400	0	1	0	1	0	1	0	1	22.92	0.69
	480	0	1	0	1	0	1	0	1	22.92	0.69
	560	0	1	0	1	0	1	0	1	22.92	0.69
ABCDEFG	288	1.79	0.97	0	1	0	1	0	1	26.79	0.62
	320	1.79	0.97	0	1	0	1	0	1	25.00	0.63
	400	1.79	0.97	0	1	0	1	0	1	25.00	0.63
	480	1.79	0.97	0	1	0	1	0	1	25.00	0.63
	560	1.79	0.97	0	1	0	1	0	1	23.21	0.67
ABCDEFGH	288	3.13	0.95	0	1	0	1	0	1	32.81	0.56
	320	3.13	0.95	0	1	0	1	0	1	28.13	0.63
	400	3.13	0.95	0	1	0	1	0	1	29.69	0.60
	480	3.13	0.95	0	1	1.56	0.98	0	1	31.25	0.57
	560	3.13	0.95	0	1	1.56	0.98	0	1	32.81	0.56

### 4.3. Comparison Waveform Before and After Embedding

One wave audio cover (drum.wav) is selected for analyzing the waveform in time domain before and after steganography, as shown in Figure 7. The first picture shows the audio carrier signal waveform; the second picture shows the audio stego signal waveform, and the last one shows the difference between the audio carrier signal waveform and the audio stego signal waveform. It can be seen from Figure 7 that the difference is so small that human beings cannot perceive the difference.



**Figure 7.** Comparison waveform of drum.wav as audio cover before and after embedding

## 5. Conclusion

Multichannel audio steganography system, based on MPEG Surround (MPS) using direct sequence spread spectrum (DSSS), has been successfully developed. MPEG Surround is used for audio steganography with multichannel audio by embedding object message into downmix signal as audio cover. The quality of the audio signal, embedded by object message, can achieve a signal-to-noise ratio (SNR) of 16.67 dB for all variations of the experiment. The object message text can be successfully extracted with a mean of bit error rate (BER) is 3.38%, and normalized cross-correlation (NC) is 95%.

## References

- [1] He J and Zhang H 2008 Digital Right Management Model Based on Cryptography and Digital Watermarking 2008 *International Conference on Computer Science and Software Engineering* vol 3 pp 656–60.
- [2] Prasetyadi G C, Mutiara A B, and Refianti R 2017 File encryption and hiding application based on advanced encryption standard (AES) and append insertion steganography method 2017 *Second International Conference on Informatics and Computing (ICIC)* pp 1–5.
- [3] Basu P N and Bhowmik T 2010 On Embedding of Text in Audio A Case of Steganography 2010 *International Conference on Recent Trends in Information, Telecommunication and Computing* pp 203–6.
- [4] Lindawati and Siburian R 2017 Steganography implementation on an Android smartphone using the LSB (a least significant bit) to MP3 and WAV audio 2017 *3rd International Conference on Wireless and Telematics (ICWT)* pp 170–4.
- [5] Binny A and Koilakuntla M 2014 Hiding Secret Information Using LSB Based Audio

- Steganography 2014 *International Conference on Soft Computing and Machine Intelligence* pp 56–9.
- [6] Kaur R, Thakur A, Saini H S and Kumar R 2015 Enhanced Steganographic Method Preserving Base Quality of Information Using LSB, Parity and Spread Spectrum Techniques 2015 *Fifth International Conference on Advanced Computing & Communication Technologies* pp 148–52.
  - [7] Nutzinger M and Wurzer J 2011 A Novel Phase Coding Technique for Steganography in Auditive Media 2011 *Sixth International Conference on Availability, Reliability and Security* pp 91–8.
  - [8] Ghasemzadeh H and Kayvanrad M H 2015 Toward a robust and secure echo steganography method based on parameters hopping 2015 *Signal Processing and Intelligent Systems Conference (SPIS)* pp 143–7.
  - [9] Nugraha R M 2011 Implementation of Direct Sequence Spread Spectrum steganography on audio data *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics* pp 1–6.
  - [10] Rezaei F, Ma T, Hempel M, Peng D and Sharif H 2013 An anti-steganographic approach for removing secret information in digital audio data hidden by spread spectrum methods 2013 *IEEE International Conference on Communications (ICC)* pp 2117–22.
  - [11] Beack S, Sung J, Seo J and Lee T 2016 MPEG surround extension technique for MPEG-H 3D Audio *ETRI J.* **38** 829–37.
  - [12] Elfriti I, Gunel B and Kondo A M 2011 Multichannel Audio Coding Based on Analysis by Synthesis *Proc. IEEE* **99** 657–70.
  - [13] Breebaart J 2007 Analysis and Synthesis of Binaural Parameters for Efficient 3D Audio Rendering in MPEG Surround 2007 *IEEE International Conference on Multimedia and Expo* pp 1878–81.
  - [14] Elfriti I, Laksono H D and Permana A K 2015 Balanced-delay filterbank for Closed-Loop Spatial Audio Coding 2015 *International Seminar on Intelligent Technology and Its Applications (ISITIA)* pp 403–8.
  - [15] Seo J, Beack S, Jang I and Jang D 2008 Development of Multichannel Sound Scene Visualization Tool with MPEG Surround Multichannel Decoder 2008 *Digest of Technical Papers - International Conference on Consumer Electronics* pp 1–2.
  - [16] Elfriti I, Muharam M and Shobirin M 2014 Distortion analysis of hierarchical mixing technique on MPEG surround standard *Proc. - ICACSYS 2014 2014 Int. Conf. Adv. Comput. Sci. Inf. Syst.* 396–400.
  - [17] Wu B and Gao L 2015 Downmix and coding of multichannel signals based on spatial correlation 2015 *8th International Congress on Image and Signal Processing (CISP)* pp 142–6.
  - [18] Gunawan T S, Rashid S A A and Kartiwi M 2017 Investigation of various algorithms on multichannel audio compression 2017 *IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)* pp 1–5.
  - [19] Elfriti I, Luthfi A and Fitrilina 2015 R-TTT module with a modified residual signal for improving multichannel audio signal accuracy 2015 *International Conference on Automation, Cognitive Science, Optics, Micro-Electro-Mechanical System, and Information Technology (ICACOMIT)* pp 207–11.
  - [20] Luthfi A, Mustaqim F and Elfriti I 2016 Studi dan Analisis Kinerja Mpeg Surround pada Bitrate 256 – 400 Kbps *J. Nas. Tek. Elektro* **5**.
  - [21] Narayana S and Prasad G 2010 Two New Approaches for Secured Image Steganography Using Cryptographic Techniques and Type Conversions *Signal Image Process. An Int. J.* **1** 60–73.