

Article

WhistleGAN for Biomimetic Underwater Acoustic Covert Communication

Yongcheol Kim ¹, Seunghwan Seol ¹ , Hojun Lee ², Geunho Park ³ and Jaehak Chung ^{1,*}

¹ Department of Electronic Engineering, Inha University, Incheon-si 22212, Republic of Korea; 22181424@inha.edu (Y.K.); koe3400@naver.com (S.S.)

² Information and Communication Engineering, Hoseo University, Cheonan-si 31499, Republic of Korea; hjlee@hoseo.edu

³ Agency of Defense Development, Changwon-si 516852, Republic of Korea; ghpark57935@add.re.kr

* Correspondence: jchung@inha.ac.kr; Tel.: +82-032-860-7421

Abstract: This paper proposes a whistle-generative adversarial network (WhistleGAN) that generates whistles for biomimetic underwater covert acoustic communication. The proposed method generates new whistles to maintain covertness by avoiding the repetitive use of the same whistles. Since the human ear perceives octave frequency such that low-frequency resolution is relatively larger than that of low frequencies, the proposed WhistleGAN uses mel filter banks to keep the fidelity in mimicking while reducing the complexity. The mean opinion score test verified that the whistles generated by the proposed method and the recorded real whistles have a similar score of 4.3, and the computer simulations proved that the bit error rate performance of the proposed method is the same as that of the real whistle.

Keywords: underwater communication; biomimetic UWA covert underwater communication; machine learning



Citation: Kim, Y.; Seol, S.; Lee, H.; Park, G.; Chung, J. WhistleGAN for Biomimetic Underwater Acoustic Covert Communication. *Electronics* **2024**, *13*, 964. <https://doi.org/10.3390/electronics13050964>

Academic Editor: Giovanni Crupi

Received: 22 January 2024

Revised: 27 February 2024

Accepted: 29 February 2024

Published: 2 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Underwater acoustic (UWA) covert communication requires a low probability of detection/interception (LPD/LPI). The conventional direct sequence spread spectrum (DSSS) method, which multiplies a communication signal by a frequency band spread code, has been widely used as a covert communication method. In UWA covert communications, however, the narrow usable bandwidth limits the large spreading factors in DSSS, and a low transmitting power reduces the communication range [1–7]. To solve this problem, researchers have studied the biomimetic UWA covert communication method that mimics and transmits the sound of marine creatures to confuse the enemy [8–17].

The biomimetic UWA covert communication mainly transmits communication signals by mimicking dolphin whistles, which propagate longer than dolphin clicks [16,17]. The conventional biomimetic UWA covert communication utilizes previously recorded dolphin whistles for communication signals. Continuous varying carrier frequency modulation (CV-CFM) in [16] divided the whistles into several slots in the time domain and allocated bits to the slots, and time frequency shift keying (TFSK) in [8] shifted the original whistles in time and frequency [8,16,17]. The combined method of CV-CFM and TFSK for transmitting bits over overlapped whistles was also investigated [17]. Since these conventional methods utilized a limited number of recorded dolphin data sets, the same data were repeatedly transmitted and the LPD/LPI performance of the biomimetic UWA covert communications decreased.

In general, two methods are used to obtain the dolphin whistles for the biomimetic UWA covert communication: one method records the real dolphin whistles by themselves and the other utilizes the published data set recorded by others. However, recording dolphins' whistles is challenging due to their movement, which can cover hundreds of

kilometers [18,19], and the quantity of published data is limited because dolphin whistles contain important marine environment information of some countries.

This paper proposes a whistle-generative adversarial network (WhistleGAN) with a Mel filter bank to create a new data set of dolphin whistles. The goal was to create new whistles that mimic the sounds of actual whistles to human listeners, while the created whistles do not have the exact same frequency contour as the recorded whistles.

Many generative adversarial networks (GAN) such as deep convolutional generative adversarial networks (DCGAN), progressive growing of GAN (ProGAN), large-scale GAN (BigGAN), and VQGAN+CLIP have been researched and present superior generation performance [20–23]. DCGAN produces high-quality images even with limited computing resources and datasets, however, it is not sufficiently good to generate high-resolution images [20]. ProGAN generates high-resolution images, yet the training process of ProGAN is complex and requires substantial computing resources [21]. BigGAN creates high-quality images through large datasets, however, it demands considerable computing time and resources for training [22]. VQGAN+CLIP generates images of various styles and resolutions, however, it has the drawback of requiring significant time and resources to train the VQGAN model [23].

In practice, acquiring a substantial database of recorded actual dolphin whistle sounds is very difficult. The purpose of this paper is to generate dolphin whistles using a low-complexity method using a relatively small amount of actual recorded dolphin whistle data. Therefore, we chose the DCGAN model for the proposed method.

The whistles generated by the proposed method provide a low computational complexity while keeping a high mimicking performance. The proposed WhistleGAN uses the Mel filter bank to convert linear frequency bands to a Mel scale because the frequency response of the human ear is perceived as octaves and lower frequencies are detected more sensitively than higher frequencies. The decreased high-frequency band resolution reduces the time-frequency data size while maintaining a high degree of whistle similarity. In addition, since the dolphin whistles exist only at specific times and frequencies in the spectrogram of the recorded data, the proposed WhistleGAN is trained with a bilateral filter that preserves an accurate contour.

To evaluate the performance of the proposed method, the complexities of the conventional generative models were compared. A mean opinion score (MOS) test was conducted to evaluate how similar the whistles generated by the proposed WhistleGAN were to the real whistle sounds. Through the MOS test, we verified that the new dolphin whistles generated by the proposed method resemble real dolphin sounds to the human ear, even though the frequency contours differ. The bit error rate (BER) performance of whistles generated by WhistleGAN were tested to evaluate the communication performance of the generated whistles. The simulation results verified that the BER performances of the proposed method are the same as actual whistles with the lower complexity of the whistle generation.

The contributions of this paper are threefold:

1. We propose the WhistleGAN to generate the new whistles with high mimicking performance with less computation by reducing the frequency data size using the Mel filter bank, which considers human hearing characteristics.
2. We conducted the MOS test to evaluate the mimicking performance of the newly created whistles. We verified that the whistle sound generated by the proposed WhistleGAN resembles real dolphin sounds.
3. We compared the BER performance of the whistles generated by the proposed WhistleGAN and the real dolphin with the conventional biomimetic UWA covert communication's modulation method. We demonstrated that the BER performance of the whistles generated by the proposed method was the same as that of the real whistles.

2. Proposed Method

2.1. Dolphin's Call Signal

The dolphin sounds used in the biomimetic UWA covert communication can be classified into clicks and whistles. Figure 1 depicts an example of a real dolphin's call signal in which the click and whistle sounds are mixed. In Figure 1, the click sounds have a short period in time with a broad frequency band and the whistle sounds have a relatively longer time and various shapes in time frequency domain.

In general, the clicks are used to localize, detect, and identify objects at close range, while the whistles are used for dolphin's social communication. The clicks are similar to pulse signals, and they have the disadvantage of having a short communication range. On the other hand, whistles have a relatively long communication range because of their long duration and variable bandwidth compared to clicks. Thus, the whistles have been mainly used in biomimetic UWA covert communication [8–17].

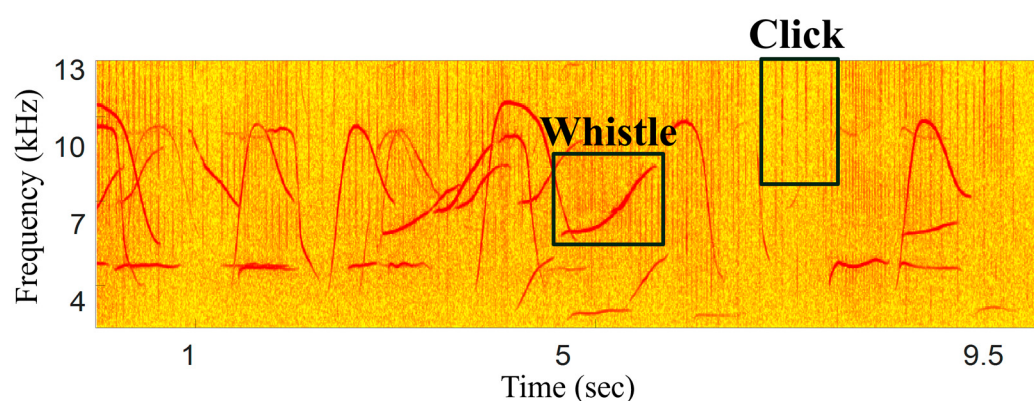


Figure 1. The dolphin's call signals.

2.2. WhistleGAN

When a whistle generative network is designed, reflecting the characteristics of the desired whistle is important. In the biomimetic UWA covert communication, the transmitted signals are eavesdropped by humans. Thus, it is necessary to generate the whistles by reflecting the characteristics of the human ear.

In this paper, we propose the WhistleGAN, which uses the Mel filter bank to generate new whistles with higher mimicking than conventional methods with a lower computation. Since human hearing is perceived in octaves, learning nonlinear data based on the human ear can produce a more imitative whistle than learning a traditional discrete Fourier transform (DFT)-based linear scale spectrogram. Learning a nonlinear scale that is log-like rather than linear in the frequency domain reduces the spectrogram data size, and whistles with a high degree of mimic (DoM) are generated by learning the human hearing characteristics.

The proposed WhistleGAN consists of a generator network (G) and a discriminator network (D) with Mel filter bank. The goal of the G is to generate fake whistles similar to the real whistles to fool the D into judging the generated whistles as the real whistles. The D takes the fake whistles generated by the G and the real whistles as input and decides whether the input is real. Since the two networks have conflicting goals, they train competitively and simultaneously to generate new whistles.

The Mel filter bank reduces the image size of the whistle spectrogram by converting a linear scale frequency to a logarithmic scale, known as the Mel scale. This processing emphasizes the low-frequency bins that correspond to human auditory sensitivity in detail, while sparsely representing high-frequency regions. The bilateral filter is used to remove noise that may occur during the GAN training process. The block diagram of the proposed WhistleGAN is shown in Figure 2.

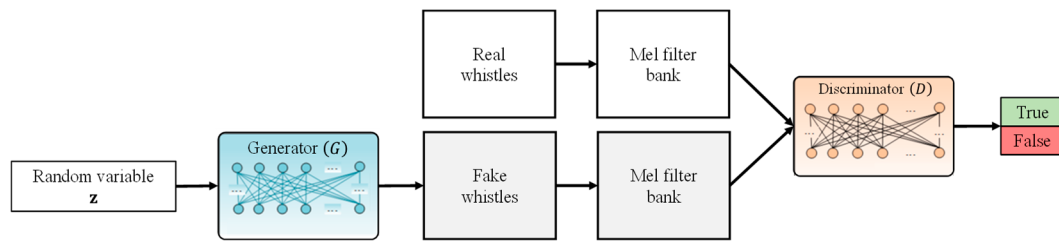


Figure 2. The block diagram of the proposed WhistleGAN.

2.3. Mel Filter Bank

This section describes using the Mel filter bank to obtain log scale frequency data from WhistleGAN. Assume that the short-time Fourier transform (STFT) of a whistle signal x is $\mathbf{X}_s$. The $\mathbf{X}_s$ of the whistle are transformed using a Mel filter bank by reducing the number of frequency bins while preserving the whistle characteristics for human hearing. The Mel filter bank is defined according to the center frequency and consists of overlapping triangular filters. The number of frequency bins in $\mathbf{X}_s$ is determined by the Mel filter bank. As the Mel filter bank number increases, the number of frequency bins in the Mel filter bank increases, resulting in an increase in the frequency domain resolution.

To design the Mel filter bank, let the linear frequency and Mel frequency be l_f and ϕ_f , respectively. ϕ_f and l_f have the relationship of $\phi_f = 2595 \times \log_{10} \left(1 + \frac{l_f}{700} \right)$. The whistle spectrogram is divided into several bands by the Mel filter bank, which consists of a series of overlapping triangular filters defined by the center frequency $l_{fc}(m)$ where $m = 1, 2, \dots, N_{Mel}$. To obtain the center frequency (Hz) of such a Mel filter bank, the inverse relationship between l_f and ϕ_f given by $l_{fc}(m) = 700 \left(10^{\phi_{fc}(m)/2595} - 1 \right)$ is used and the Mel filter bank $H(m, k)$ where $k = 0, 1, \dots, N_{DFT} - 1$ is obtained as,

$$H(m, k) = \begin{cases} 0, & \text{for } l_f(k) < l_{fc}(m-1) \\ \frac{l_f(k) - l_{fc}(m-1)}{l_{fc}(m) - l_{fc}(m-1)}, & \text{for } l_{fc}(m-1) \leq l_f(k) \leq l_{fc}(m) \\ \frac{l_{fc}(m) - l_f(k)}{l_{fc}(m) - l_{fc}(m+1)}, & \text{for } l_{fc}(m) < l_f(k) \leq l_{fc}(m+1) \\ 0, & \text{for } l_f(k) \geq l_{fc}(m+1). \end{cases} \quad (1)$$

The Mel filter bank matrix $\mathbf{H}$ consists of rows of the number of Mel filter banks and columns of the size of DFT. To obtain Mel whistles ($\mathbf{X}_{mel}$), $\mathbf{X}_s$ is multiplied by the $\mathbf{H}$ and taken logarithmically.

$$\mathbf{X}_{mel} = \log \{ \mathbf{H} \cdot |\mathbf{X}_s| \}, \quad (2)$$

where $\cdot$ denotes inner product.

The whistles used for the input of the WhistleGAN are located in a specific time-frequency region. The proposed WhistleGAN trains using a bilateral filter that considers the distance between the target pixels of the whistle image and the surrounding pixels, as well as the difference in pixel values. This allows WhistleGAN to preserve the whistle's shape while removing noise. The whistle image with the bilateral filter ($\mathbf{X}_{bi}$) applied to $\mathbf{X}_{mel}$ is calculated through adaptive convolutions of the pixel values' similarity denoted as $f_r(\cdot)$ and the proximity of pixel positions represented by $g_s(\cdot)$.

Let W_p be the normalized term, μ be the position of the current pixel, and Ω be the window centered in μ . Then, $\mu_v \in \Omega$ denotes the position of the neighboring pixel. When $\mathbf{X}_{mel}(\mu_v)$ and $\mathbf{X}_{mel}(\mu)$ are assumed as the pixel's value at the corresponding position in the Mel whistle, ($\mathbf{X}_{bi}$) is attained by,

$$f_r(\mu_v, \mu) = \exp \left(\frac{-\|\mu_v - \mu\|^2}{2\sigma_d^2} \right), \quad (3)$$

$$g_s(\mu_v, \mu) = \exp\left(\frac{-\|\mu_v - \mu\|^2}{2\sigma_r^2}\right), \quad (4)$$

$$W_p = \sum_{\mu_v \in \Omega} f_r(\|\mathbf{X}_{mel}(\mu_v) - \mathbf{X}_{mel}(\mu)\|) g_s(\|\mu_v - \mu\|), \quad (5)$$

$$\mathbf{X}_{bi}(\mu) = \frac{1}{W_p} \sum_{\mu_v \in \Omega} \mathbf{X}_{mel}(\mu_v) f_r(\|\mathbf{X}_{mel}(\mu_v) - \mathbf{X}_{mel}(\mu)\|) g_s(\|\mu_v - \mu\|), \quad (6)$$

where σ_d and σ_r denote the smoothing parameters. In this paper, an amplified whistle ($\mathbf{X}_e$) is learned from the D of the WhistleGAN by amplifying $\mathbf{X}_{bi}$ generated by bilateral filter by α .

$$\mathbf{X}_e = \alpha \times \mathbf{X}_{bi}. \quad (7)$$

In the G , a random variable $\mathbf{z}(\mathbb{R}^{1 \times Z})$ is used to generate a new whistle. The D learns the features of a whistle obtained from the Mel filter bank in the WhistleGAN ($\mathbf{X}_e$) and a whistle with no processing ($\mathbf{X}_s$). The G trains to satisfy $D(G(\mathbf{z})) = 1$, while the D trains to satisfy $D(G(\mathbf{z})) = 0$ through a competitive process. The trained WhistleGAN generates the whistle contours with the same probability distribution. The competition process between G and D of the proposed WhistleGAN can be attained as,

$$\min_G \max_D V(D, G) = E_{\mathbf{X}_s \sim P_{data}(\mathbf{X}_s)} [\log D(\mathbf{X}_s)] + E_{\mathbf{X}_e \sim P_{data}(\mathbf{X}_e)} [\log D(\mathbf{X}_e)] + E_{\mathbf{z} \sim P_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (8)$$

where E denotes the expectation. $P_{data}(\mathbf{X}_s)$, $P_{data}(\mathbf{X}_e)$ and $P_z(\mathbf{z})$ denote the probability distributions of $\mathbf{X}_s$, $\mathbf{X}_e$, and $\mathbf{z}$, respectively.

2.4. Structure of WhistleGAN

The G in the WhistleGAN uses fractional-strided convolution instead of a fully connected layer to generate whistles. To solve the problem of gradient vanishing/exploding, batch normalization (BN) is applied in all layers except the output layer. The output layer uses the tanh activation function, and the other uses the ReLU activation function.

The D of the WhistleGAN uses strided convolution instead of a fully connected layer to reduce the number of features in the input whistle image and determine true/false. The D utilizes batch normalization in all layers except the input and output layers and uses the sigmoid activation function to determine true/false in the output layer. In the remaining layers, the LeakyReLU activation function is employed, which introduces a slope of alpha to allow negative values unlike ReLU and thus enables the learning of weights to be influenced.

Since the WhistleGAN contains the Mel filter bank, each network learns the input data with different values and sizes than the conventional DCGAN. Thus, the kernel and stride sizes are optimized through iterative research. The specific structure of the G of the proposed WhistleGAN is shown in Table 1 and the specific structure of the D of the WhistleGAN is shown in Table 2.

Table 1. The specific structure of the WhistleGAN's generator network.

	Kernel	Stride	BN	Activation Function
Input	4×4	1	Y	ReLU
Conv1	4×4	2	Y	ReLU
Conv2	4×4	2	Y	ReLU
Conv3	4×4	2	Y	ReLU
Conv4	4×4	2	Y	ReLU
Output	4×4	2	N	Tanh

Table 2. The specific structure of the WhistleGAN’s discriminator network.

	Kernel	Stride	BN	Activation Function
Input	4×4	2	N	LeakyReLU
Conv1	4×4	2	Y	LeakyReLU
Conv2	4×4	2	Y	LeakyReLU
Conv3	4×4	2	Y	LeakyReLU
Conv4	4×4	2	Y	LeakyReLU
Conv5	4×4	2	Y	LeakyReLU
Output	2×2	2	N	Sigmoid

3. Analysis of Experimental Results

In this section, the conventional generative models were compared for the selection of the low complexity generative model. To evaluate the proposed method, the whistle generation performance, the similarity between generated and real whistles, and the BER performance for the proposed method were compared by visualization, human evaluation, and computer simulation, respectively.

3.1. Complexity Comparison

We compared the complexity of ProGAN, BigGAN, VQGAN+CLIP, and DCGAN. Table 3 presents the number of nodes for each technique according to two input sizes of $128 \times 128 \times 3$ and $256 \times 256 \times 3$. DCGAN exhibits 14 times less complexity than ProGAN, 50 times less complexity than BigGAN, and 5 times less complexity than VQGAN+CLIP, respectively. Since one of the purposes of the proposed method is reduction of the computational complexity, the performance of the proposed method was compared with that of the conventional DCGAN.

Table 3. The number of nodes in generation models according to input size.

Method	$128 \times 128 \times 3$	$256 \times 256 \times 3$
DCGAN [20]	10 million	30 million
ProGAN [21]	140 million	420 million
BigGAN [22]	500 million	1500 million
VQGAN+CLIP [23]	50 million	150 million

3.2. Whistle Generation Results

For the training of the proposed WhistleGAN, the whistles from the “Watkins Marine Mammal Sound Database (WMMSD)”, and “voice in the sea and discovery of sound in the sea” were used [24,25]. The data of about 239 single whistles were increased to 10,000 whistles by changing the time and frequency and used as the training set.

To compare the computational complexity and accuracy between the proposed WhistleGAN and the conventional methods, the DCGAN (256) and the proposed WhistleGAN were trained, and DCGAN (128) was also trained with the same whistles of the same size for a fair comparison. The whistle inputs for DCGAN (256) and DCGAN (128) were set to $256 \times 256 \times 3$ and $128 \times 128 \times 3$, respectively, according to frequency, time, and color. Note that the proposed WhistleGAN with a 128 Mel filter bank was trained by reduced whistle image to $128 \times 128 \times 3$, which was four times smaller than the DCGAN (256) input image size. The frequency scale of the input whistles of the proposed method was transformed into a logarithmic Mel scale with equal intervals from the linear scale.

For training the proposed WhistleGAN and the conventional DCGAN, the input whistles were normalized between -1 and 1 . The WhistleGAN and DCGAN were implemented on Pytorch 1.12.0 and trained using RTX 4090 (NVIDIA, SCL, Santa Clara, CA, USA). Both methods used the same learning parameters for a fair comparison in Table 4. The bench test results of the computational complexity of DCGAN and the proposed WhistleGAN are shown in Table 5.

Table 4. The hyperparameters of the DCGAN and the proposed WhistleGAN.

Hyperparameter	Value
Batch size	128
Epochs	2000
Optimizer	Adam
Learning rate	2×10^{-4}
Adam's beta	0.5, 0.999
LeakyReLU's alpha	0.2

Table 5. The bench test results of the DCGAN and the proposed WhistleGAN.

	DCGAN (256)	DCGAN (128)	WhistleGAN
Average batch time (ms)	15.65	6.60	9.61
Average throughput (images/second)	4088.29	15,431.82	13,324.42
Estimated total size (MB)	2730.67	810.58	1450.58

In Table 5, the average batch time of the proposed WhistleGAN was 1.63 times faster than that of the conventional DCGAN (256). The number of images processed per second of the proposed method was 3.25 times higher. The total parameter size of the proposed method was 1.88 times smaller. Thus, the proposed WhistleGAN has 47% less computation than the conventional DCGAN (256) using smaller-sized whistles, even with additional filters and the learning process. On the other hand, the conventional DCGAN (128) with the same size whistle input used in the proposed method required less computational complexity than the proposed WhistleGAN because the conventional DCGAN (128) did not use the bilateral filter.

Figures 3 and 4 depict examples of whistles generated by the DCGAN (256) and DCGAN (128), and Figure 5 displays examples of whistles generated by the proposed WhistleGAN. Figure 3 shows the whistle of size $256 \times 256 \times 3$, and Figures 4 and 5 show the whistle of size $128 \times 128 \times 3$. In Figure 5, the y-axis denotes the Mel frequency scale, while others utilize linear scale. The whistles generated by the DCGAN and the proposed WhistleGAN are noise-free and generated various whistle contours such as flat, up-chirp, down-chirp, and fluctuations, etc. [26].

Since the method generated new whistle patterns, the conventional correlation assessment between real and generated whistles cannot be applied. Therefore, the qualitative assessment of the similarity was tested by humans using the MOS test in the next subsection.

3.3. DoM Assessment

The MOS test utilizes the BS1284 international telecommunication union standard, which is used to evaluate the degree of distortion of call sounds in wireless communications [27]. The MOS test serves as a metric that averages the perceived degree based on the auditory senses of multiple individuals. It requires the participation of at least 20 individuals and employs a 5-point scale for scoring. The evaluation table is shown in Table 6. The higher the average MOS test score, the higher the DoM.

For the MOS test comparison, 31 participants rated the real dolphin sounds, the sounds generated by the DCGAN (256) and DCGAN (128), and the proposed WhistleGAN, in a randomized order. Shure SRH 840 headphones were used for listening. To achieve a similar sound quality, the marine ambient noise extracted from real whistle audio files were added to the generated whistles. Figure 6 shows examples of the real whistles used in the experiments, the generated whistles by DCGAN, and the generated whistles by the proposed WhistleGAN. Figure 6 shows some whistles generated by the DCGAN and WhistleGAN that are similar to the original whistles.

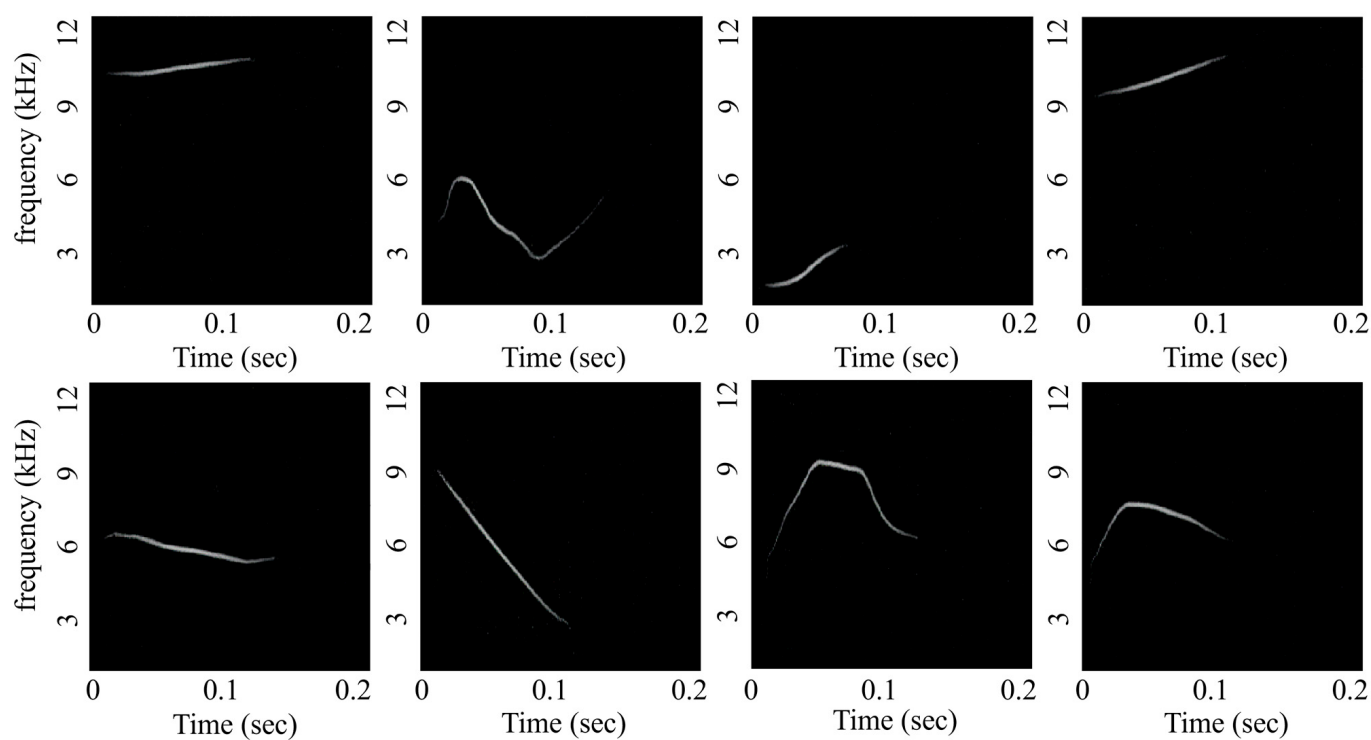


Figure 3. The whistle examples generated by the DCGAN (256).

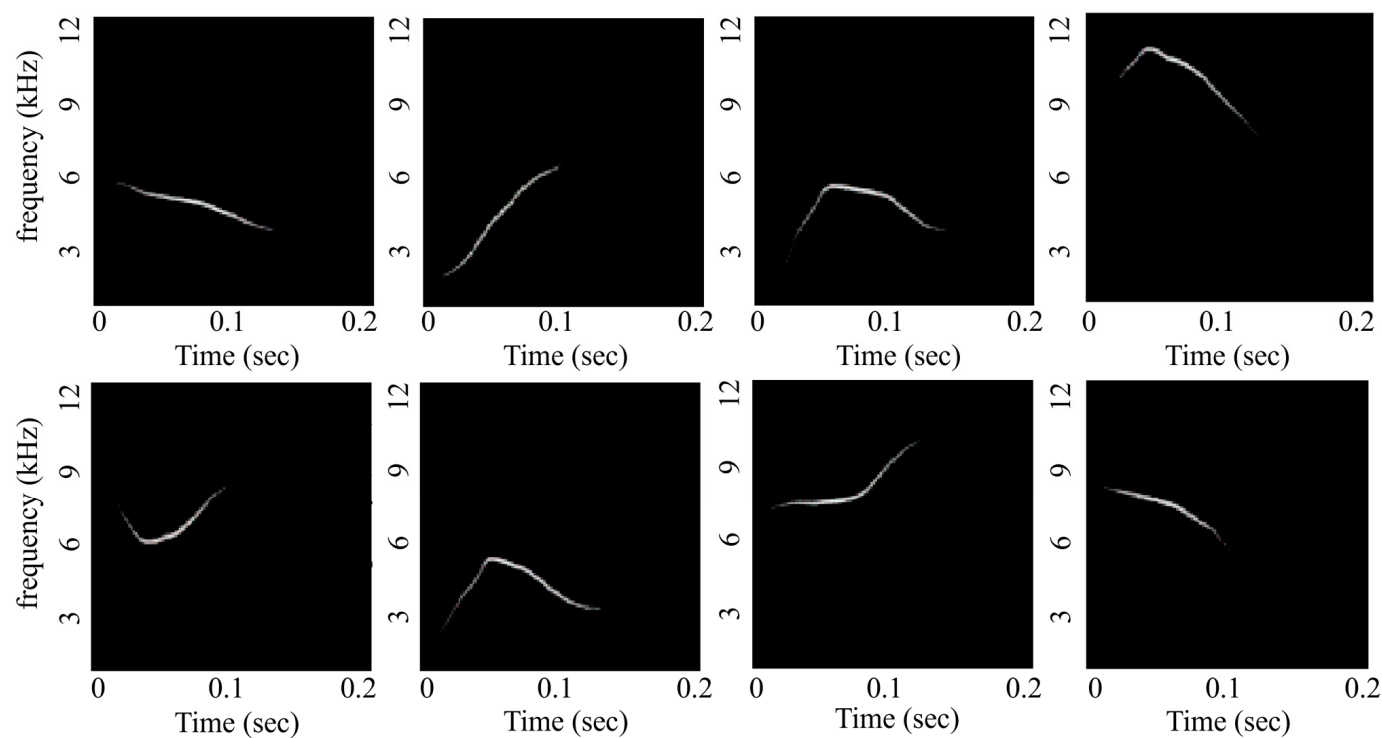


Figure 4. The whistle examples generated by the DCGAN (128).

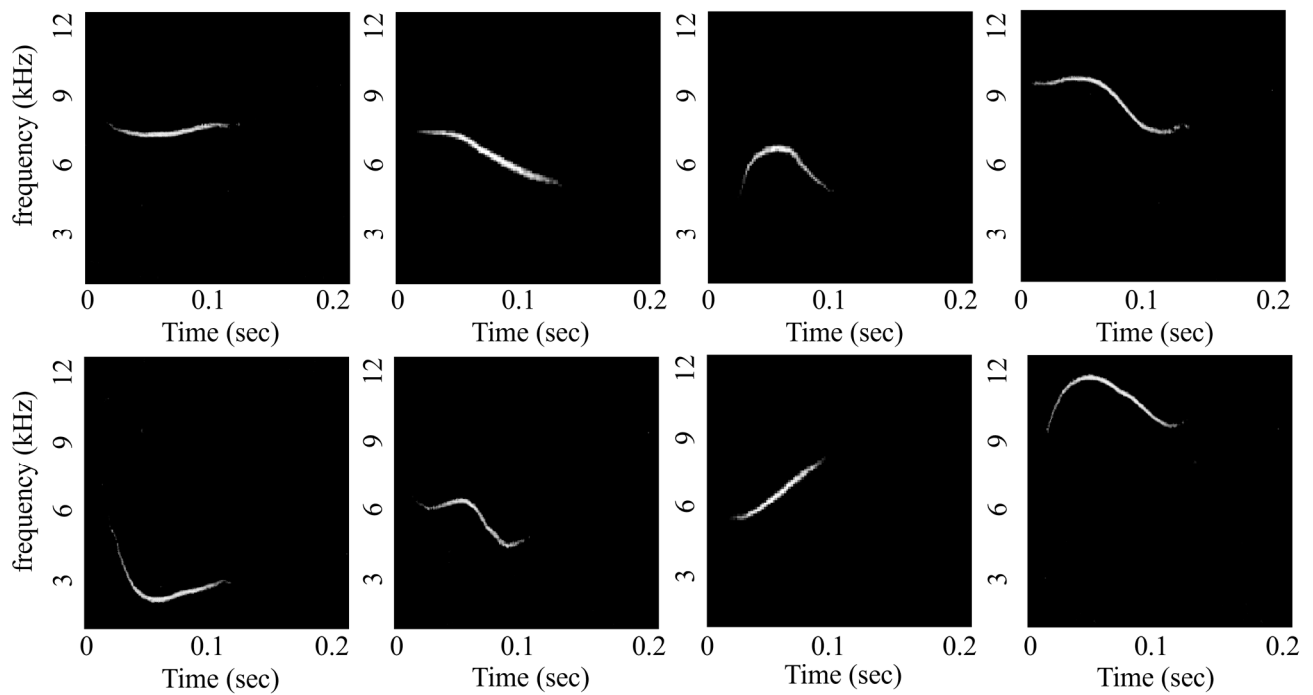


Figure 5. The whistle examples generated by the proposed WhistleGAN.

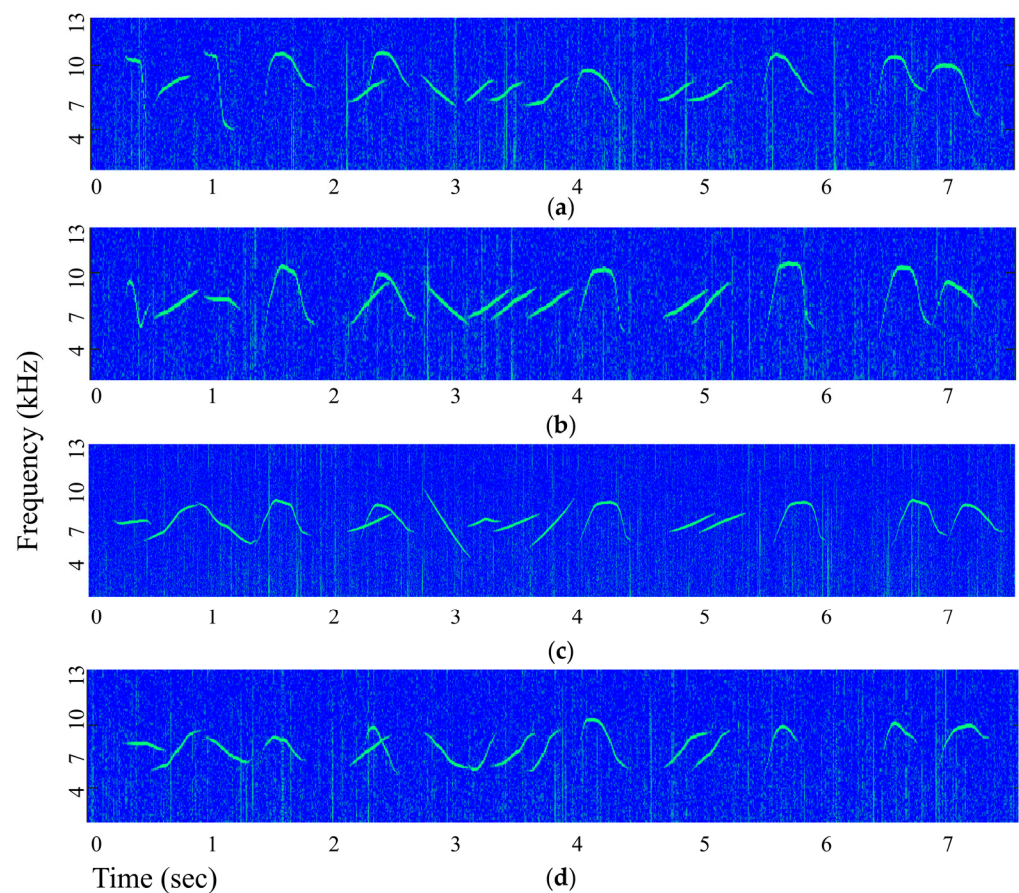


Figure 6. The spectrograms of (a) real whistles; (b) generated whistles by the DCGAN (256); (c) generated whistles by the DCGAN (128) (d) generated whistles by the proposed WhistleGAN.

Table 6. The MOS test scoring criteria.

Score	5	4	3	2	1
Opinion	Same	Very similar	Similar	Slightly different	Different

Table 7 depicts the results of the MOS test. In Table 7, the real dolphin sound scored 4.3, the sound generated by the DCGAN (256) scored 4.2, the sound generated by the DCGAN (128) scored 4.0, and the sound generated by the proposed WhistleGAN scored about 4.3. The proposed WhistleGAN and the DCGAN (256) had 100% and 98% similarity to the original signal, respectively, but DCGAN (128) had 93% similarity, which is less similar to the original whistles.

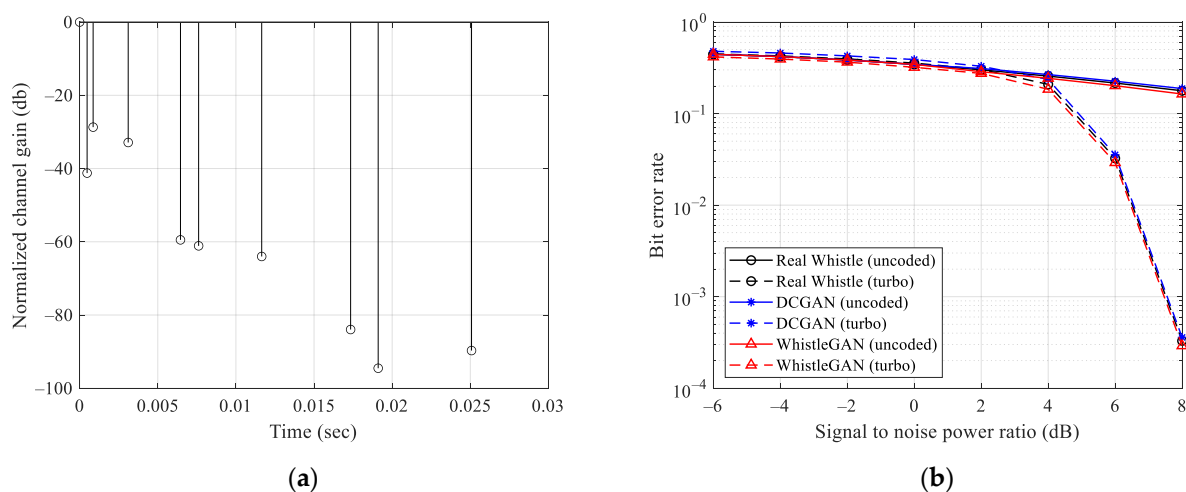
Table 7. The MOS test results.

Scheme	Real dolphin	DCGAN (256)	DCGAN (128)	WhistleGAN
Score	4.3	4.2	4.0	4.3

Through the MOS test, we verified that the proposed WhistleGAN generated new whistles with a high DoM and low complexity by learning the probability distribution of the original whistles.

3.4. Communication Performance Assessment

In this subsection, the BER of the biomimetic UWA communication using the real dolphin whistles and the generated whistles are compared. The DCGAN (256) and the proposed WhistleGAN were used to generate the whistles. DCGAN (128) was not used for the comparison because the DCGAN (128) has low DoM. The modulation methods used for the communication were CV-CFM, orthogonal frequency division multiplexing (OFDM), and hybrid orthogonal whistle modulation (HOWM). The data rates of the three techniques were set to 50 bps [16,17]. The modulation parameters of all methods were the same as in previous studies and a symbol bandwidth was 200 Hz and the forward error correction was a 1/3 turbo code [16,17]. The computer simulations were performed on three UWA channels attained from the West Sea, South Sea, and East Sea of Korea. The channel impulse responses (CIRs) and BER results of each UWA channel are shown in Figures 7–9.

**Figure 7.** Cont.

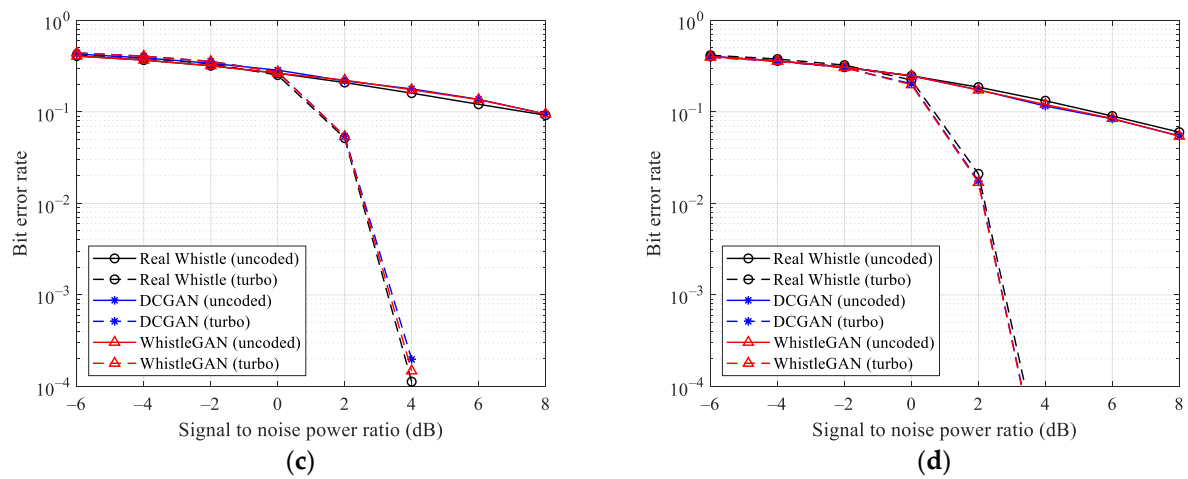


Figure 7. (a) The CIR of the West Sea of South Korea; the BER results of (b) CV-CFM; (c) OFDM; and (d) HOWM.

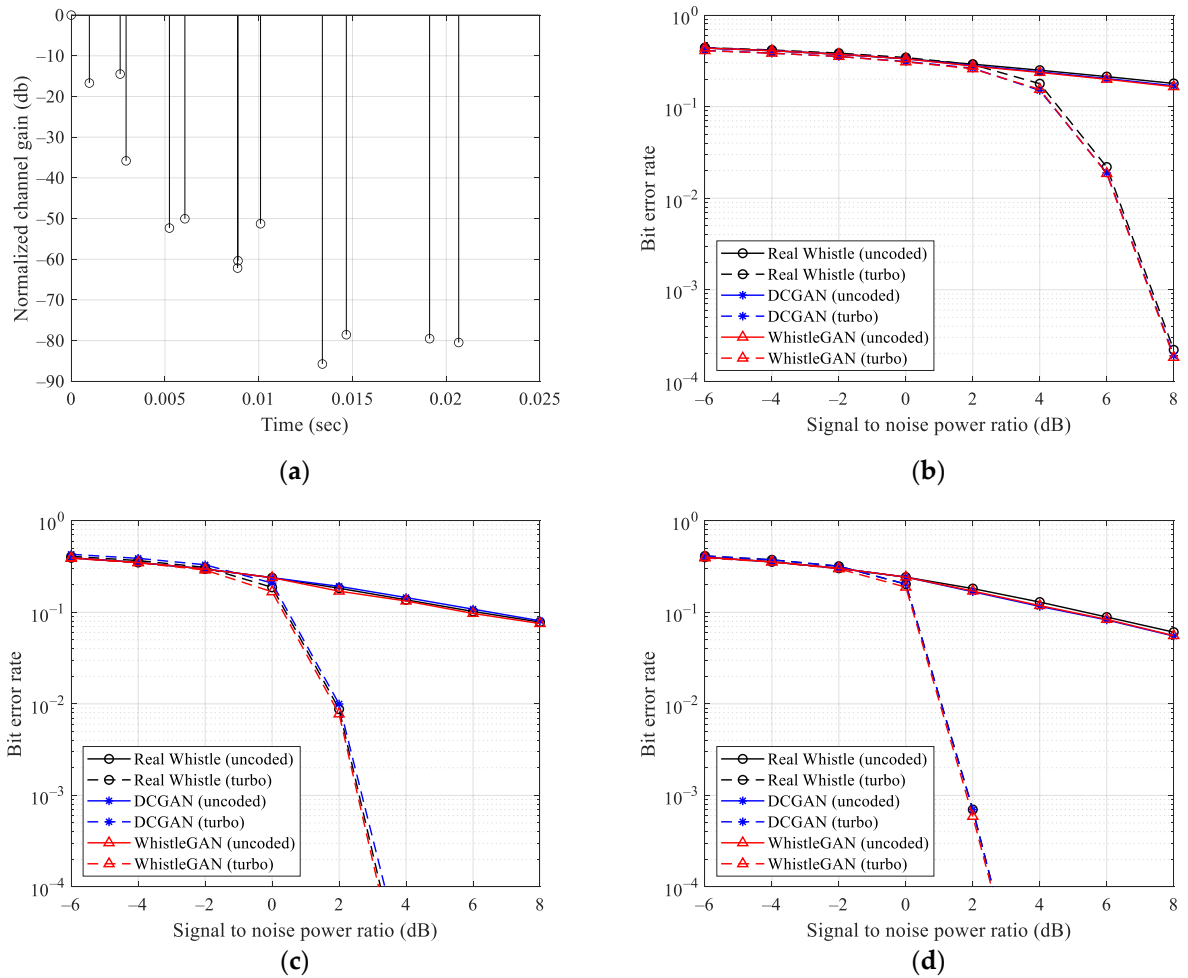


Figure 8. (a) The CIR of the South Sea of South Korea; and the BER results of (b) CV-CFM; (c) OFDM; (d) HOWM.

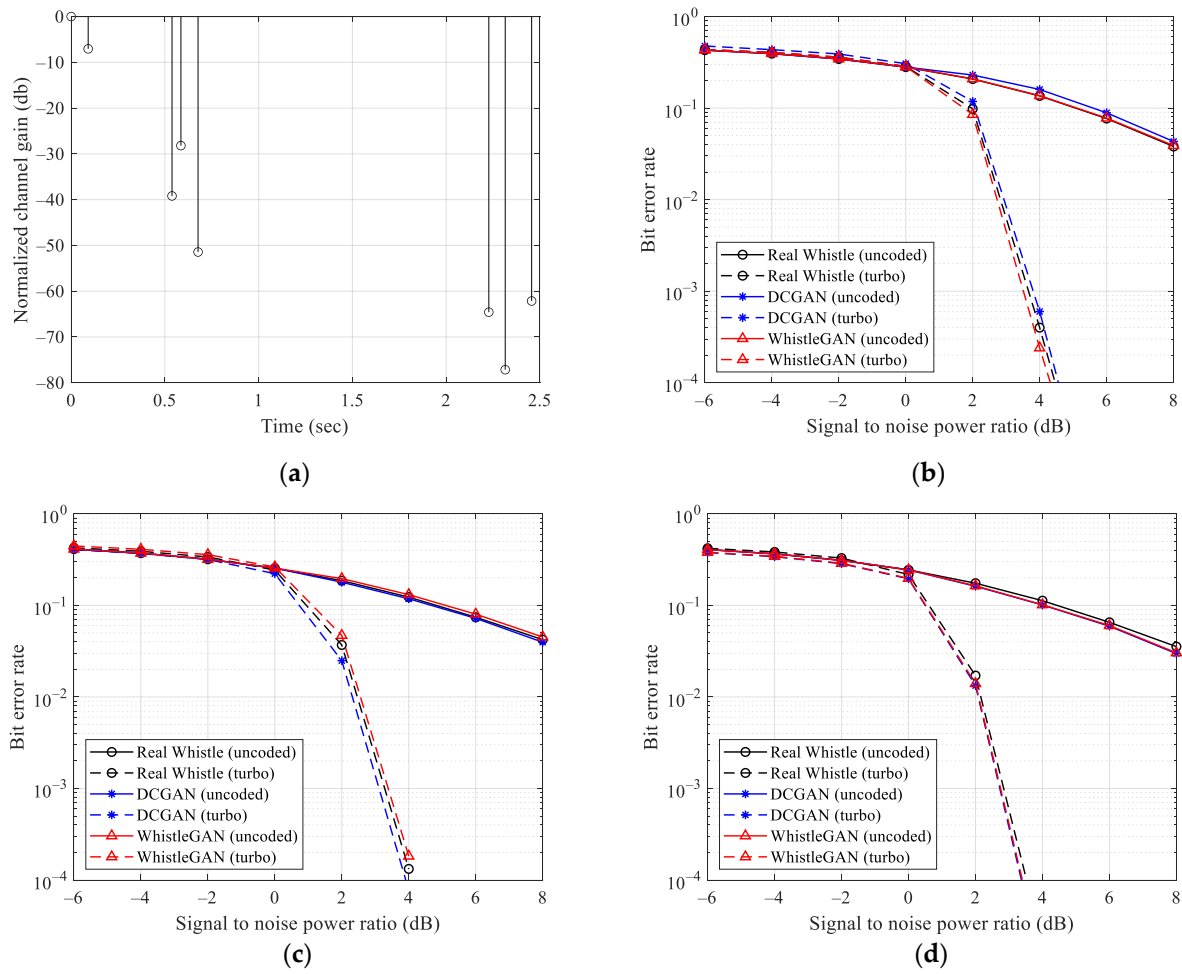


Figure 9. (a) The channel impulse response of the East Sea of South Korea; and the BER results of (b) CV-CFM; (c) OFDM; (d) HOWM.

Figures 7a–d, 8a–d and 9a–d depict the CIRs, the BERs of the CV-CFM, BERs of the OFDM, and the BERs of the HOWM, respectively. The black line, the blue line, and the red line denote the BERs with the real whistles, with the whistles by the DCGAN, and with the proposed WhistleGAN. The solid lines denote the uncoded BER and the dashed lines denote the coded BER.

Figures 7–9 demonstrate the similar BER performances for the the real dolphin and the generated whistles by the proposed WhistleGAN through the computer simulations of the three conventional biomimetic UWA covert communication methods and the three different UWA channel models. Therefore, We verified that the proposed WhistleGAN generates a variety of dolphin whistles with a high DoM and similar BER performance to real dolphin sounds using a lower complexity than the conventional DCGAN (256).

4. Conclusions

This paper proposed the WhistleGAN that generated whistles with different contours to those of real dolphin whistlers. The proposed WhistleGAN used the Mel filter bank, which modified the frequency scale for human ear characteristics resulting in a low computational complexity and a high DoM. The MOS tests were conducted on WhistleGAN, the DCGAN, and real dolphin whistles. The proposed WhistleGAN had 7% better DOM when the computation complexity was similar to DCGAN, and 2% better DOM when the computation complexity was lower by 50%. The BER performances were compared to real whistles, whistles generated by the proposed WhistleGAN and whistles generated by the conventional DCGAN through three different underwater channels with three different

biomimetic UWA covert communication methods. The proposed scheme exhibited the same BER as real whistles.

Author Contributions: Conceptualization, Y.K.; methodology, Y.K.; software, Y.K. and H.L.; validation, Y.K. and H.L.; formal analysis, Y.K. and H.L.; investigation, Y.K.; resources, J.C.; data curation, G.P. and S.S.; writing—original draft preparation, Y.K.; writing—review and editing, J.C.; visualization, Y.K.; supervision, J.C.; project administration, Y.K. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Agency for Defense Development by the Korean Government (UI237029DG).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, T.C.; Yang, W.B. Low probability of detection underwater acoustic communications using direct-sequence spread spectrum. *J. Acoust. Soc. Am.* **2008**, *124*, 3632–3647. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ling, J.; He, H.; Li, J.; Roberts, W. Covert underwater acoustic communications. *J. Acoust. Soc. Am.* **2010**, *128*, 2898–2909. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Shu, X.; Wang, J.; Wang, H.; Yang, X. Chaotic direct sequence spread spectrum for secure underwater acoustic communication. *J. Appl. Acoust.* **2016**, *104*, 57–66. [\[CrossRef\]](#)
4. Diamant, R.; Lampe, L. Low probability of detection for underwater acoustic communication: A Review. *IEEE Access Underw. Wirel. Commun. Netw.* **2018**, *104*, 19099–19112. [\[CrossRef\]](#)
5. Qu, F.; Qin, X.; Yang, L.; Yang, T.C. Spread-spectrum method using multiple sequences for underwater acoustic communications. *J. Ocean Eng.* **2018**, *43*, 1215–1225. [\[CrossRef\]](#)
6. Schmidt, J.H. Using fast frequency hopping technique to improve reliability of underwater communication system. *J. Appl. Sci.* **2019**, *10*, 1172. [\[CrossRef\]](#)
7. Ko, S.J.; Kim, W.J. Robust frame synchronization algorithm in time varying underwater acoustic communication channel. *J. Acoust. Soc. Korea* **2020**, *39*, 8–15.
8. Lee, H.J.; Ahn, J.M.; Kim, Y.C.; Lee, S.K.; Chung, J.H. A biomimetic communication method based on time shift using dolphin whistle. *J. Acoust. Soc. Korea* **2019**, *38*, 580–586.
9. Song, L.; Gang, Q.; Asim, I. Covert underwater acoustic communication using dolphin sounds. *J. Acoust. Soc. Am.* **2013**, *133*, EL300–EL306.
10. Xiao, H.; Jingwei, Y.; Pengyu, D.; Xiao, Z. Experimental demonstration of underwater acoustic communication using bionic signals. *Appl. Acoust.* **2014**, *78*, 7–10.
11. Ahmad, E.; Meng, Z.; Tolga, M.D.; Antonia, P.S. An underwater acoustic communication scheme exploiting biological sounds. *Wirel. Commun. Mob. Comput.* **2016**, *16*, 2194–2211.
12. Songzuo, L.; Mengjia, W.; Tianlong, M.; Gang, Q.; Muhammad, B. Covert underwater communication by camouflaging sea piling sounds. *Appl. Acoust.* **2018**, *142*, 29–35.
13. Muhammad, B.; Songzuo, L.; Gang, Q.; Lei, W.; Yan, T. Bionic Morse coding mimicking humpback whale song for covert underwater communication. *Appl. Sci.* **2019**, *10*, 186.
14. Gang, Q.; Tianlong, M.; Songzuo, L.; Muhammad, B. A frequency hopping pattern inspired bionic underwater acoustic communication. *Phys. Commun.* **2021**, *46*, 101288.
15. Jiajia, J.; Chunyue, L.; Xianquan, W.; Zhongbo, S.; Xiao, F.; Fajie, D. Covert underwater communication based on combined encoding of diverse time-frequency characteristics of sperm whale clicks. *Appl. Acoust.* **2021**, *171*, 07660.
16. Ahn, J.M.; Lee, H.J.; Kim, Y.C.; Lee, S.K.; Chung, J.H. Mimicking dolphin whistles with continuously varying carrier frequency modulation for covert underwater acoustic communication. *Jpn. J. Appl. Phys.* **2019**, *58*, SGGF05. [\[CrossRef\]](#)
17. Kim, Y.C.; Lee, H.J.; Seol, S.H.; Park, B.G.; Chung, J.H. Underwater Biomimetic Covert Acoustic Communications Mimicking Multiple Dolphin Whistles. *Electronics* **2023**, *12*, 3999. [\[CrossRef\]](#)
18. Hellen, B.; Paul, T. Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. *J. Anim. Ecol.* **2006**, *75*, 456–465.
19. Dinis, A.; Alves, F.; Nicolau, C.; Ribeiro, C.; Kaufmann, M.; Canadas, A.; Freitas, L. Bottlenose dolphin *Tursiops truncatus* group dynamics, site fidelity, residency and movement patterns in the Madeira Archipelago (North-East Atlantic). *Afr. J. Mar. Sci.* **2016**, *38*, 151–160. [\[CrossRef\]](#)
20. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In Proceedings of the ICLR 2016, San Juan, Puerto Rico, 2–4 May 2016.
21. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In Proceedings of the ICLR 2018, Vancouver, BC, Canada, 1–3 May 2018.

22. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In Proceedings of the ICLR 2019, New Orleans, LA, USA, 6–9 May 2019.
23. Crowson, K.; Biderman, S.; Kornis, D.; Stander, D.; Hallahan, E.; Castricato, L.; Raff, E. VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. In Proceedings of the ECCV 2022, Tel Aviv, Israel, 25–27 October 2022.
24. Watkins Marine Mammal Sound Database. Available online: <https://go.who.edu/marine-mammal-sounds> (accessed on 12 March 2021.).
25. Discovery of Sound in the Sea. Available online: <https://dosits.org/> (accessed on 12 March 2021.).
26. Chmelnitsky, E.G.; Ferguson, S.H. Beluga whale, *Delphinapterus leucas*, vocalizations from the Churchill River. *J. Acoust. Soc. Am.* **2012**, *131*, 4821–4835. [[CrossRef](#)] [[PubMed](#)]
27. BS1284; General Methods for the Subjective Assessment of Sound Quality. International Telecommunication Union: Geneva, Switzerland, 2003.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.