한국음향학회지 제43권 제1호 pp. 9~18 (2024)
The Journal of the Acoustical Society of Korea Vol.43, No.1 (2024)
https://doi.org/10.7776/ASK.2024.43.1.009

pISSN : 1225-4428
eISSN : 2287-3775

실데이터 기반 능동 소나 신호 합성 방법론

Real data-based active sonar signal synthesis method

김윤수,[1] 김주호,[2] 석종원,[1†] 홍정표[1†]

(Yunsu Kim,[1] Juho Kim,[2] Jongwon Seok,[1†] and Jungpyo Hong[1†])

[1]창원대학교 정보통신공학과, [2]국방과학연구소 해양기술연구원 소나체계개발단

(Received May 19, 2023; revised August 2, 2023; accepted December 5, 2023)

초 록: 최근 수중표적의 저소음화와 해상교통량의 증가로 인한 주변 소음의 증가로 능동 소나 시스템의 중요성이 증대되고 있다. 하지만 신호의 다중 경로를 통한 전파, 다양한 클러터와 주변 소음 및 잔향 등으로 인한 반향신호의 낮은 신호대잡음비는 능동 소나를 통한 수중 표적 식별을 어렵게 만든다. 최근 수중 표적 식별 시스템의 성능을 향상 시키기 위해 머신러닝 혹은 딥러닝과 같은 데이터 기반의 방법을 적용시키려는 시도가 있지만, 소나 데이터셋의 특성 상 훈련에 충분한 데이터를 모으는 것이 어렵다. 부족한 능동 소나 데이터를 보완하기 위해 수학적 모델링에 기반한 방법이 주로 활용되어오고 있다. 그러나 수학적 모델링에 기반한 방법론은 복잡한 수중 현상을 정확하게 모의하는 데에는 한계가 있다. 따라서 본 논문에서는 심층 신경망 기반의 소나 신호 합성 기법을 제안한다. 제안하는 방법은 인공지능 모델을 소나 신호 합성 분야에 적용하기 위해, 음성 합성 분야에서 주로 사용되는 타코트론 모델의 주요 모듈인 주의도 기반의 인코더 및 디코더를 소나 신호에 적절하게 수정하였다. 실제 해상 환경에 모의 표적기를 배치해 수집한 데이터셋을 사용하여 제안하는 모델을 훈련시킴으로써 보다 실제 신호와 유사한 신호를 합성해낼 수 있게 된다. 제안된 방법의 성능을 검증하기 위해, 합성된 음파 신호의 스펙트럼을 직접 분석을 진행하여 비교하였으며, 이를 바탕으로 오디오 품질 인지적 평가(Perceptual Quality of Audio Quality, PEAQ)인지적 성능 검사를 실시하여 총 4개의 서로 다른 환경에서 생성된 반사 신호들에 대해 원본과 비교해 그 차이가 최소 −2.3이내의 높은 성적을 보여주었다. 이는 본 논문에서 제안한 방법으로 생성한 능동 소나 신호가 보다 실제 신호에 근사한다는 것을 입증한다.

핵심용어: 능동 소나 신호, 수중 음향 신호 처리, 신호 합성, 딥러닝

ABSTRACT: The importance of active sonar systems is emerging due to the quietness of underwater targets and the increase in ambient noise due to the increase in maritime traffic. However, the low signal-to-noise ratio of the echo signal due to multipath propagation of the signal, various clutter, ambient noise and reverberation makes it difficult to identify underwater targets using active sonar. Attempts have been made to apply data-based methods such as machine learning or deep learning to improve the performance of underwater target recognition systems, but it is difficult to collect enough data for training due to the nature of sonar datasets. Methods based on mathematical modeling have been mainly used to compensate for insufficient active sonar data. However, methodologies based on mathematical modeling have limitations in accurately simulating complex underwater phenomena. Therefore, in this paper, we propose a sonar signal synthesis method based on a deep neural network. In order to apply the neural network model to the field of sonar signal synthesis, the proposed method appropriately corrects the attention-based encoder and decoder to the sonar signal, which is the main module of the Tacotron model mainly used in the field of speech synthesis. It is possible to synthesize a signal more similar to the actual signal by training the proposed model using the dataset collected by arranging a simulated target in an actual marine environment. In order to verify the performance of the proposed method, Perceptual evaluation of audio quality test was conducted and within score difference −2.3 was shown compared to actual signal in a total of four different environments. These results prove that the active sonar signal generated by the proposed method approximates the actual signal.

Keywords: Active sonar signal, Underwater acoustic signal processing, Signal synthesis, Deep learning

PACS numbers: 43.30.Vh, 43.72.Ja

†**Corresponding author:** Jongwon Seok (jwseok@changwon.ac.kr), Jungpyo Hong (hansin@changwon.ac.kr)
Department of Information and Communication Engineering, Changwon National University, 20 Changwondaehak-ro, Uichang-gu, Changwon, Gyeongnam 51140, Republic of Korea
(Tel: 82-55-213-3836, 82-55-213-3838, Fax: 82-55-213-3839)

I. 서 론

소나란 수중에서 목표로하는 물체의 존재, 위치, 특성을 파악하는 장비 를 말한다.[1] 그 종류로는 수동 소나와 능동 소나 두 가지가 있으며, 수동 소나는 수신기만이 존재하는 탐지기로서 함선의 엔진과 프로펠러와 같은 물체 자체가 생성하게 되는 진동을 탐지한다. 이는 설계가 간편하며 구축에 사용되는 비용이 적지만 동물 및 다른 함선의 신호까지 모두 여과 없이 수신하여 원하는 신호만을 구분하기 위해서는 방대한 양의 데이터를 필요로 한다. 반면에 능동 소나는 송신기가 임의의 신호를 방사하여 원하는 물체에 반사되어 돌아온 신호를 수신기가 받아들인다. 방사한 신호는 사전에 설정된 주파수 특성을 가지고 있어 탐지가 용이하며, 돌아오는 시간을 측정하여 거리를 측정하는 것 역시 가능하다. 능동 소나의 이와 같은 특성과 함께 최근 잠수함의 정숙화와 해상교통량의 증가로 능동소나에 대한 중요성이 증대되고 있다.[2,3]

능동 소나의 표적 인식 성능을 높이기 위해 기계학습과 같은 인공지능 모델을 접목 시키려는 연구 및 시도가 되어오고 있으며, 이에 따라 학습하기에 충분한 양질의 데이터셋을 필요로 하게 된다. 그러나 수집 비용이 높은 소나 신호의 특성 상 학습 데이터셋의 구축이 어렵게 되는데, 이에 실제 해상 환경에서 장비를 구동시켜 얻은 것과 유사한 소나 신호를 생성해내는 것이 요구되고 있다.[3]

소나 데이터의 모의 신호 생성을 위해 다양한 연구 기관 및 기업이 해양환경 모델링 기반의 시뮬레이션 방법을 제시했는데, 그중 하나가 North Atlantic Treaty Organization(NATO)해저 연구 센터에 의해 제공된 시뮬레이션 모듈으로 송신기가 방사한 신호를 통계학에 기반한 계산을 통해 표적과 접촉 상태를 모의한다.[4] 또한 실제 바다환경 데이터셋에서 볼 수 있는 센서 간 표적 페이딩 효과를 제공하여 더욱 사실적인 신호를 생성해낸다. 하지만 소나 방정식을 단순화시키는 데에 있어 불가피하게 실제 환경에서 수집된 데이터와는 차이가 있게 되었다. 이에 추가적으로 잔향 및 단순화한 해양환경을 반영한 Multi-Everything Sonar Simulation(MESS)가 발표되었다.[5] 하지만 이 역시 기존 소나 방정식에 단순화시킨 해양 환경 파라미터들을 추가하였기 때문에 실제 데이터에 가깝게 구현해내지 못했다. 또한, SIMulator Of Non-acoustics- and Acoustics(SIMONA)는 접촉 상태, 잔향 뿐만아니라 표적의 형태, 다중경로 페이딩, 파형의 종류 등을 반영한 신호를 생성한다.[6] 전체 시뮬레이션을 위해 빔포밍 및 정합 필터 모듈에 필수로 입력되어야하는 잔향의 계산이 보다 사실적인 데이터 생성에 주요한 역할을 하게된다. 따라서 양방향 능동 소나에 한해 실시간으로 잔향음을 생성하는 연구 역시 진행되었다.[7] 하지만 위와 같은 수학적 모델링 기반 방법론들은 방대하며 복잡한 실제 수중 환경을 정확히 모사하는데 한계를 보인다.

한편, 심층 신경망의 빠른 발전에 따라 단순한 텍스트 정보로부터 가변적인 길이의 복잡한 시계열 신호를 생성하는 기술에 많은 관심 및 연구가 진행되었다[8-12]. WaveNet[8]은 2016년에 제안된 음성 신호 합성 모델로서, 심층 신경망이 오디오 샘플 생성에 두드러지는 성능을 보인 대표적인 사례이다. 다만 입력으로 자연어 텍스트가 아닌 원하는 음성의 언어적 특징이 담겨있는 멜 스펙트로그램을 입력으로 사용하는 일종의 보코더의 역할로서까지만 사용된다는 한계가 있다. DeepVoice는[9] 기존 Text-To-Speech(TTS) 파이프라인을 심층 신경망으로 대체한 방법이다. 그러나 학습이 단대단으로 되지 않는다는 단점이 존재한다. 이후 합성 성능을 올리기 위해 인코더-디코더 구조의 모델이 제안되었으며, 사전 훈련된 은닉 마르코프 모델을 사용해 중요도를 계산하여 보코더 파라미터를 예측한다.[10] 단대단 학습을 가능하게 하기 위해서 설계된 Char2Wav[11] 모델 역시 있지만, 여전히 보코더 파라미터를 예측한다는 점에서 추가적인 전처리를 필요로 한다. 타코트론은[12] 2017년에 공개된 TTS 모델로서, 자연어 텍스트에서 음성 데이터의 선형 스펙트럼을 한번에 훈련에 성공한 단대단 모델이다. 인코더와 디코더, 그리고 주의도 3개의 서브모듈로 구성하여 상용 어플리케이션에 활용이 가능할 정도로 높은 생성 성능을 보여주어 차후 제안된 TTS 모델들의 기본 구조로 많이 사용된다.

따라서 본 연구에서는 능동소나 신호 생성을 위해 기존의 수학적 모델링 기반의 신호 생성기법과는 완

전히 다른 접근 방법인 심층 신경망 기반의 능동 소나 신호 합성 기법을 제안한다. 이러한 목표를 이루기 위해 음성합성 분야의 대표적인 방법인 타코트론 모델을 소나 신호 생성에 적용하였고, 타코트론을 적용하는 과정에서 음성합성과 소나신호 합성의 입출력의 근본적인 차이를 반영하기 위해 필요한 주요 모듈을 소나신호 합성에 적합하게 변형하였다.

II. 관련 연구

2.1 하이라이트 모델 기반 능동 소나 표적신호 합성[13]

고주파를 사용하는 능동 소나의 반사 신호는 물체의 하이라이트의 공간적 분포에 의해 특징화된 내부의 여러 등가 산란과 함께 물체 표현의 정반사에 의해 만들어진다. 다상태 능동 소나 모델링은 수중 표적에 맞아 반사된 신호를 모의하는 것을 의미한다. 정상 상태에서 수중 물체에게 펄스 신호를 방사하였을 때, 선체, 매질, 구조적 특징, 입사파의 주파수 및 펄스 너비와 같은 요소들로 인해 다양한 형태의 반사 신호가 발생하게된다. 소나 신호를 모의하는 것은 이러한 반사의 과정에서 발생할 수 있는 모든 것을 고려하는 것이다. 표적에 부딪히는 점을 각각 모두 반사 추적 알고리즘에 입력하는 것은 무한대의 경우의 수를 가지기 때문에 일련의 점으로서 표적을 모의하는 하이라이트 개념을 도입하게 된다.

장거리에서 수중 표적은 하나의 하이라이트로부터 발생한 단일점으로 표현된다. 하지만 단거리에서 표적은 시간과 각도에 따라 달라지는 분포 특성을 가질 수 있기 때문에 하이라이트의 분포를 적절하게 표현할 필요가 있다. 반사된 신호에 영향을 미치는 특정 위치에 부착된 하이라이트의 입사각에 따라 변화하는 표적의 표면에 불연속적인 인지한 후 잠수함을 표적으로 가정한 회전타원체 하이라이트 개념을 사용한다. 이와 같은 모의 실험 방법론은 단순하지만 높은 환경 근사성으로 범용적으로 사용되는 모델이다. 하지만 표적 및 해양환경을 단순화 시키는 과정에서 생기는 실제 신호와의 차이점, 특히 배경 잡음을 모델링함에 있어 두드러지는 차이를 보여주어 기계학습모델의 훈련 데이터셋으로 사용하기에는 부족하다.

2.2 텍스트 기반 음성 합성 모델 : 타코트론[12]

타코트론은 2017년 구글에서 공개한 TTS 모델로서, TTS 시스템의 일부인 보코더의 역할을 대체하도록 하는 기존의 방법론들에서 텍스트를 입력하면 출력으로 모델링된 음성 파형이 직접적으로 나오도록 설계된 심층 신경망 모델이다. 인코더-디코더의 구조를 사용하고 있으며, 주의도를 핵심 구조로 차용한다. <텍스트, 오디오> 쌍이 각각 모델의 입력과 출력으로 구성되어있다. 텍스트는 지정 언어의 자연어를 원 데이터 그대로 입력되게하며, 출력되는 선형 스펙트로그램을 후처리를 통해 wav 오디오 파일로서 저장된다. 인코더는 자연어 텍스트 데이터를 받아 입력받은 텍스트 시퀀스의 의미를 가장 잘 나타내는 벡터인 일종의 텍스트 임베딩을 출력해준다. 임베딩된 텍스트 벡터는 디코더가 순차적으로 오디오 샘플들을 생성할 때 참고할 정보로 사용된다. 주의도 기법은 디코더가 오디오 시퀀스를 생성할때 사용하는 텍스트 임베딩 벡터의 중요도를 매 시간단계마다 결정해주는 중개자 역할을 한다. 기존 순환신경망 기반 seq-seq 모델에서 입력 문장의 길이가 길어질수록 디코딩에 있어 중요도가 높은 단어라도 문장의 초반에 위치하면 정보 자체가 서서히 사라지는 기울기 소실 문제를 완화시키며 생성 성능을 더욱 높여주었다. 타코트론은 높은 합성 성능으로 단대단 TTS 모델의 초석이 되었다.

2.3 타코트론 기반 능동 소나 신호 합성 모델[14]

입력되는 자연어 텍스트를 그 의미를 인코딩하여 가변적인 길이의 오디오 파일로 출력해주는 TTS 모델 타코트론의 특징 및 구조를 차용하여 능동 소나 신호 합성 모델이 제안되었다.

Fig. 1은 TTS 모델과 해당 연구에서 제안하는 소나 신호 합성 모델의 입출력 구조를 보여준다. 음성 신호 생성을 목적으로 하는 TTS 모델은 입력되는 자연어 문장을 한글의 경우 초성, 중성, 종성 단위의 기호로 순서대로 나누어 그 순서에 대응하는 음성 신호

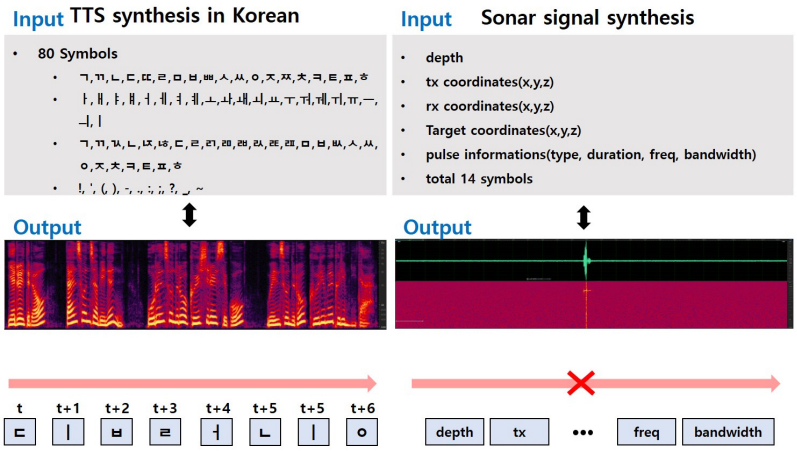

Fig. 1. (Color available online) Input and output structure of Text-To-Speech and sonar synthesis model.

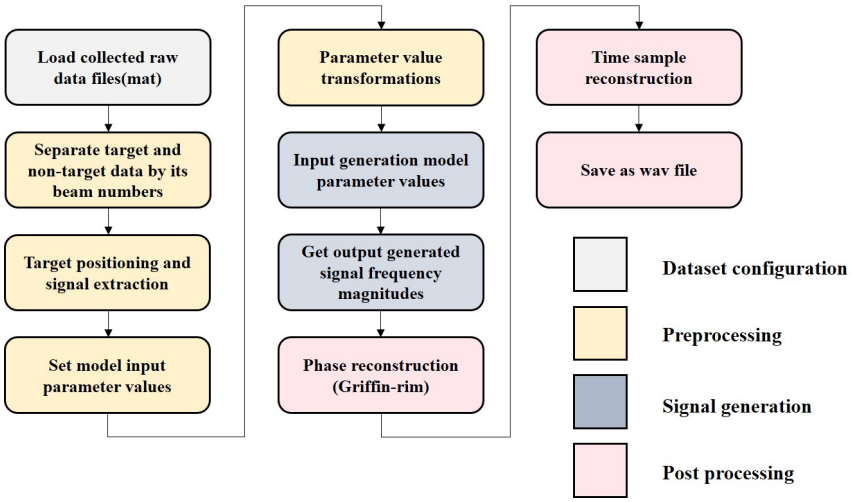

Fig. 2. (Color available online) Proposed system structure.

를 출력으로 합성하게 된다. 즉, 각 기호의 순서가 출력 신호에 영향을 미치게 되는 구조이지만, 소나 신호 데이터셋의 구조는 순서의 관계없이 해양환경 변수를 입력하게 된다. 이후 합성 모델은 입력받은 정보를 종합적으로 판단하여 이에 대응하는 전체 신호를 출력하는 구조이다. 따라서 기존의 TTS 모델과 입력되는 정보들의 순서에 관계없이 신호 합성에 필요한 정보를 판단해야 한다는 관점에서 모델 구조의 변경은 불가피하다. 따라서 입력으로 사용되는 수심, 송·수신기 및 표적 좌표, 펄스 정보까지 총 14개의 정보를 수치화해 음성 대상 모델과 구분되는 임베딩 계층을 도입하여 인코딩 및 디코딩 될 수 있도록 하였다. 하지만 해당 시스템이 합성하는 신호는 모의기에서 출력된 신호로서, 실효성을 가지고 있지 않다.

III. 제안한 방법

3.1 전체 시스템 구조

본 논문에서 제안하는 다상태 능동 소나 합성기의 전체 구조는 Fig. 2와 같다. 전체 시스템은 크게 데이터셋 구성, 전처리, 신호 생성 및 후처리까지 총 4단계로 나뉜다. 데이터셋은 사전에 기술한 바와 같이 다양한 해상환경에 실제 모의 표적기를 배치하여 수집이 진행되었다. 이렇게 수집된 데이터셋은 전처리부를 거쳐 모델에 입력될 데이터로 변환을 거치게되며, 이렇게 처리된 데이터는 심층 신경망 모델을 통해 모의되어 생성된 선형 스펙트로그램을 출력한다. 생성된 모의 신호는 후처리를 거치며 최종 파형 데이터로 저장되게 된다. Fig. 3은 본 논문에서 제안하

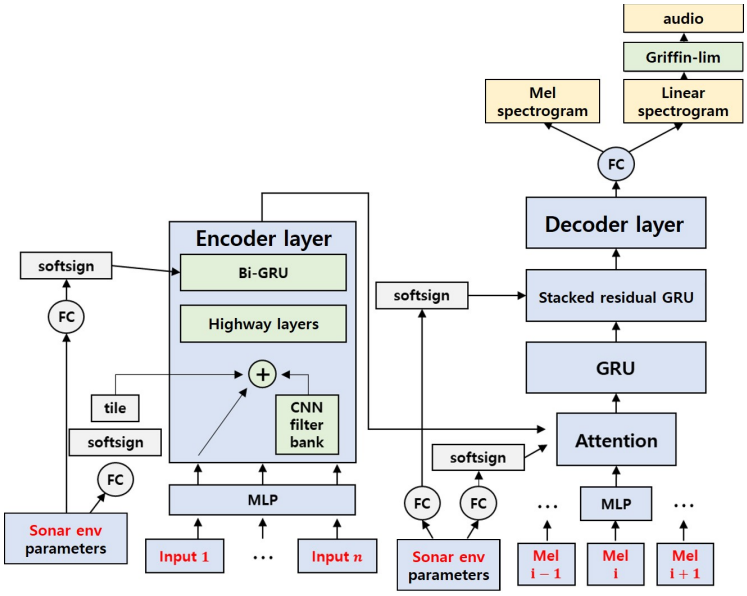

Fig. 3. (Color available online) Model structure of the proposed sonar signal synthesis.

는 시스템의 신호 생성 부분을 담당하는 심층 신경망의 구조이다. 전체적인 구조는 타코트론의 구조를 차용하였다. 모델의 입력으로는 데이터셋의 구성에 사용된 파라미터 값들을 해양환경 종류, 표적/비표적 여부, 방사 신호 종류, 진행 시간 순으로 각각 (0,1) 범위의 실수값으로 정규화 후 제공된다. 이후 인코더 네트워크를 거쳐 신호 모의에 필요한 파라미터 정보들에서 선형 스펙트로그램을 생성하는데에 필요한 정보로 추출 및 변환하는 과정을 거치게 된다.

3.2 타코트론 기반 다상태 소나 생성 모델

3.2.1 소나 환경 파라미터 임베딩 계층

기존 TTS 모델의 경우는 자연어 텍스트를 벡터로 변환하는 과정에서 토큰화를 진행한다. 사전에 설정된 단어 사전 내에서 해당 단어가 등장하는 순서를 원핫벡터로 변환한뒤, 해당 벡터를 텍스트 임베딩 계층을 통해 신경망이 문장 내에서 단어의 의미를 스스로 판단할 수 있도록 한다. Count Vectorization (C-V),[15] bag-of-words[16] 그리고 Term Frequency-Inverse Document Frequency(TF-IDF)[17]와 같이 기존에 설정되어있는 단어 의미 추출 알고리즘보다 각 문제에 특화된 의미을 추론한다는 점에서 더욱 효과적인 접근법이다.[18]

그러나 본 논문에서 소나 신호 생성에 사용되는 심층 신경망의 입력은 수심, 펄스 정보, 송수신기 및 표적의 좌표와 같은 환경 파라미터 값을 일련의 수치 벡터로 표현한다. 사전 내에서 존재하는 단어의 의미만을 추론하는 텍스트 임베딩 계층과는 달리 소나 환경 파라미터 임베딩 계층은 연속적인 수치로서 그 경우의 수는 무한대가 될 수 있다. 또한 원핫벡터 특성상 1인 부분을 제외하면 공간만을 차지하고 있는 무의미한 0값들이 원소로서 채워져 있지만, 소나 환경 파라미터 벡터는 더욱 밀집한 구조로 원소별로 고유한 의미를 내포하고 있다. 따라서 각 파라미터에 따라 개별적으로 의미를 추론할 수 있도록 가중치 벡터를 설계할 필요가 있다.

3.3 표적 마스크 $L1$ 손실함수

음성 합성 모델의 최적화를 위해서는 적절한 비용함수의 설계가 필수적이다. 모델의 출력한 음성 신호의 선형 스펙트로그램과 실제 신호를 비교한 것을 비용함수로 설계하기위해 기존의 방법은 Eq. (1)과 같은 평균 절대 오차(Mean Absolute Error, MAE)를 사용하였다.

$$L_{total} = \frac{1}{T}\frac{1}{N}\sum_{t=0}^{T}\sum_{i=0}^{N}|o_t^i - \hat{o_t^i}|. \quad (1)$$

시간 정보, 즉 전체 프레임의 수는 T, 모델이 출력하는 t번째 프레임의 i번째 주파수 스펙트럼 계수 $\hat{o_t^i}$, 실제 음성 신호의 주파수 스펙트럼 계수 o_t^i로 설정하였다. o_t^i와 $\hat{o_t^i}$의 $L1$ 거리를 전체 프레임과 계수를 거쳐 평균을 취한 값을 손실값으로 사용하였고 이는 평균제곱오차(Mean Squared Error, MSE)를 사용하는 것보다 더욱 좋은 성능을 보였다.[12]

하지만 앞서 설명한 바와 같이 소나 신호 합성 모델은 디코딩 단계에서 시간 정보가 효과적으로 전달되지 못한다. 위치기반 디코딩을 사용하여 시간적인 정보를 디코더에게 제공하더라도 보조의 역할일 뿐 근본적인 해결책은 되지 못한다. 또한 소나 신호의 특성상 특정 시점의 표적 신호를 제외하고는 전체 시간에 걸쳐 배경 잡음 혹은 클러터 신호가 대부분을 차지하고 있기때문에 표적이 있는 부분에 원본과의 차이를 줄이는 것에 조금 더 집중을 할수 있도록 비용함수를 설계할 필요가 있다. 본 논문에서 이와 같은 문제를 해결하기 위해 표적 마스크평균 절대오차를 제안한다. 표적 신호의 주파수 계수의 크기는 주로 배경 잡음의 크기보다 큰 점을 활용하여 Eq. (3)과 같이 표적으로 추정되는 곳에는 1, 표적이 없는 곳으로 추정되는 곳은 0인 마스크 M을 계산한다. 이를 신경망 모델의 출력값과 원 신호의 주파수 스펙트럼에 원소별 곱을 통해, 표적이 있는 곳끼리만 에너지를 비교하도록 하는 $L_{maskedLinear}$를 전체 비용함수 L_{total}에 추가하였다.

$$\mu = \frac{1}{N}\sum_{i=0}^{N} s_t^i. \tag{2}$$

$$M = \begin{cases} 0, & s_t^i \le 2\mu \\ 1, & s_t^i > 2\mu \end{cases}. \tag{3}$$

$$L_{maskedLinear} = \frac{1}{T}\frac{1}{N}\sum_{t=0}^{T}\sum_{i=0}^{N} |M \odot s_t^i - M \odot \hat{s_t^i}|. \tag{4}$$

$$L_{total} = L_{linear} + L_{maskedLinear}. \tag{5}$$

Eqs. (4)와 (5)는 제안하는 표적 마스크 손실함수와 신경망의 전체 손실함수를 도식화한 것이다. 여기서 $\odot$은 아다마르 곱을 의미한다.

IV. 실험 환경 및 결과

4.1 모의 표적기를 이용한 해상실험 데이터

본 논문에서 소개하는 능동 소나 신호 합성 시스템은 모의 신호를 훈련 데이터셋으로 사용하는 선행연구 방법[14]과 달리 실제 해양환경에서 수집된 데이터셋을 사용한다. 총 4가지의 다양한 해양환경에서 임의의 거리에 세워놓은 모의 표적을 향해 신호를 방사한다. 이 표적에 반사되어 다시 수신되기까지 수집한 30 s에서 60 s 사이의 1차원 신호를 하나의 데이터 파일로서 저장하여 전체 데이터셋을 구성한다. 수집이 진행된 해양 환경은 각 실험 해역, 방사 신호, 수신 신호 세기, 신호 종류, 중심 주파수 등 다양한 변수를 변경해가며 Case 1-4까지 총 4가지로 나뉘게 된다. 표적의 종류는 연속파(Continuous Wave, CW)와 선형 주파수 변조(Linear Frequency Modulation, LFM)를 사용하였으며, 실험에 사용된 표적 및 비표적 데이터들을 각 케이스에 따라 그 세부사항을 Table 1에 명시해 놓았다.

4.2 실험 환경

전체 데이터셋에서 사용된 소나 신호 파일들은 30 s대에서 시작하여 60 s까지 가변적이며, 심층 신경망 모델의 입력으로 사용하기에 상당히 긴 시간의 샘플들을 담고 있다. 이 중 표적을 맞고 돌아온 신호만을 가져와 약 4 s 가량의 작은 부분으로 나눈 뒤 프

Table 1. Total dataset configuration.

Case number	Signal strength (SNR)	Target type	Target presence	# of data
Case 1	very strong	CW, LFM	Target	1,751
			Non target	17,791
Case 2	very weak	CW, LFM	Target	311
			Non target	2,644
Case 3	strong	LFM	Target	648
			Non target	27,452
Case 4	weak	CW, LFM	Target	135
			Non target	540

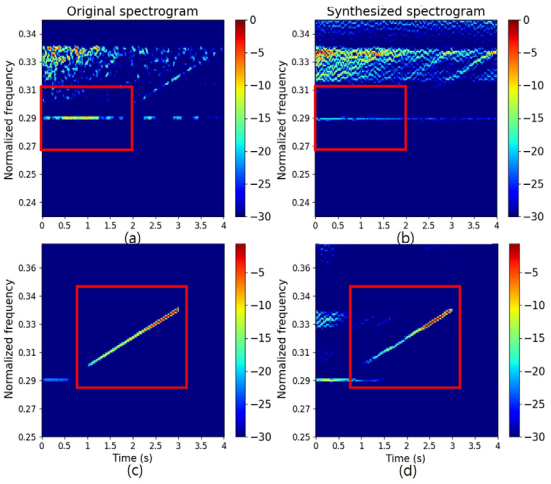

Fig. 4. (Color available online) Sonar data spectrogram comparison (Case 1). (a) Original CW, (b) synthesized CW, (c) original LFM, (d) synthesized CW.

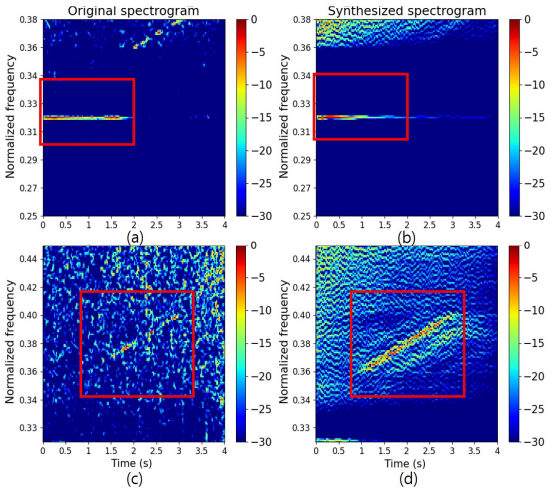

Fig. 5. (Color available online) Sonar data sample spectrogram comparison (Case 4). (a) Original CW, (b) synthesized CW, (c) original LFM, (d) synthesized CW.

레임 단위로 나누어 훈련용 입력 데이터로 사용하게 된다. 다음 기술되는 내용은 실험 과정에서 사용된 모든 변수들에 대해 기술한다. 오디오 처리, 심층 신경망 훈련 2가지로 나누어 구분하며, 실험 변수 명, 수치, 변수 설명 순으로 구성되어 있다.

a) 오디오 처리 변수

- num_mels : 80, 멜 스펙트로그램을 구하기 위한 멜 필터의 개수
- num_freq : 1,025, 주파수 계수의 개수
- frame_length_ms : 50 [ms], 프레임의 길이
- frame_shift_ms : 12.5 [ms], 프레임간 이동되는 길이

b) 모델 훈련 파라미터

- parameter embedding dimension : 256, 파라미터 임베딩 차원, 인코더의 입력 차원과 동일
- Attention type : Bahdanau attention, 주의도 종류
- Attention dimension : 256, 주의도 계산 차원
- Decoder input dimension : 256, 디코더의 입력 차원
- Decoder output dimension : num_freq, num_mels, 디코더의 출력 차원, 모델의 최종 출력 차원을 의미

4.3 실험 결과

Figs. 4와 5는 실제 능동 소나 신호와 동일한 입력 변수로 합성된 신호 샘플의 스펙트로그램을 보여준다. 붉게 표시된 영역은 표적이 현재 위치하는 대역을 의미한다. CW 신호는 주로 0 s ~ 1 s까지 부근 영역에 올바르게 신호를 합성하는 것을 확인할 수 있으며, LFM 신호는 2 s 부근에 신호의 중심부가 위치하게 된다. 실제 신호와 합성된 신호가 동일한 위치에 대채로 유사하게 생성되는 것을 확인할 수 있다.

Fig. 6은 실제 소나 신호와 합성된 신호의 선형 스펙트로그램 주파수 계수 에너지 오차를 3차원의 형태로 보여준다. 마찬가지로 일부 구간을 제외한 대부분의 구간이 적은 절대오차 값을 보였다. 다만, Fig. 10에서 표적 신호가 존재하는 대역에 일부 오차가 큰 부분은 사용된 타코트론 모델의 손실함수에 의한 영향으로 관측된다. 해당 심층 신경망 모델의 손실함수는 MAE를 사용하는데, MSE는 원본과의 전체적인 차이를 줄이는 방향으로 최적화를 진행해 나가는 반면, MAE는 중요하다고 판단되는 특징은 더욱 부각시키며, 이외의 특징은 억제하는 방향으로 학습을 진행하게 된다. 신호 값의 등락이 큰 시계열 음파 신호의 특성 상 본 논문의 모델에서 역시 평균 MAE를 사용하였으나 표적의 특징만을 지나치게 부각시켜 그림과 같은 에너지의 차이를 보이게 된다. 또한, 배경잡음의 경우 스펙트로그램상에 불규칙적으로 분포되어 있는 원본 능동 소나 신호와는 달리 규칙적인 형태의 잡음을 생성한다. 이는 보편적인 학습 구조의 심층 신경망에서 발생하게 되는 문제이

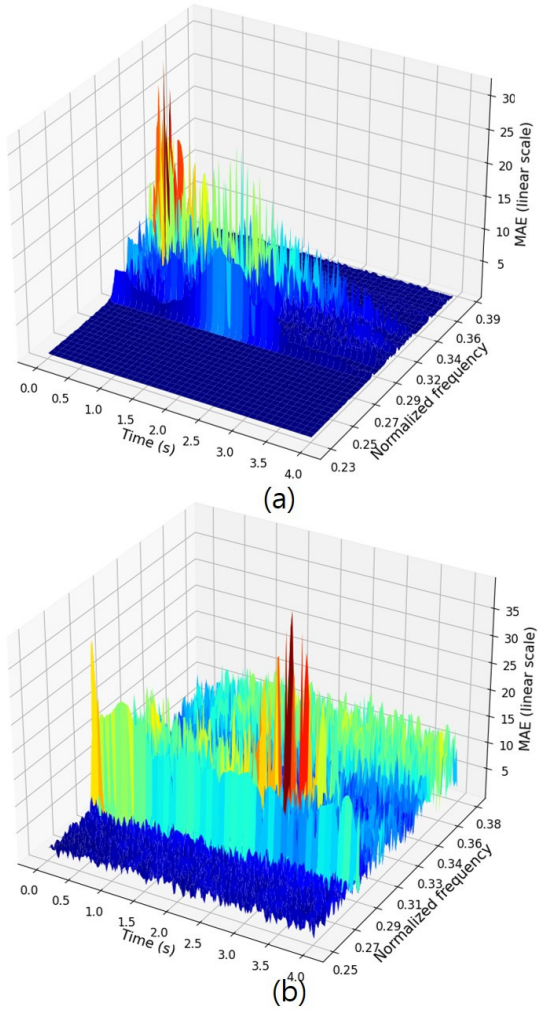

Fig. 6. (Color available online) MAE between original and synthesized active sonar signals on the target band. (a) CW signal, (b) LFM signal.

다. MAE 혹은 MSE와 같은 오차 기반의 신경망은 주어진 데이터 내에서 공통되는 특징을 찾아 이를 바탕으로 신호를 합성하게 되는데, 이러한 규칙성이 결과에 반영되어 인공적인 잡음을 초래하게 된다.[19] 또한 실제 신호에서 나타나는 극적인 값의 변화를 표현하기에 역시 일반적인 신경망이 낮은 성능을 보여준다. 따라서 보다 실제와 유사하게 생성하도록 적대적 생성 학습법[20]을 차용하여 사실성을 더욱 부각시키는 방법론이 시도되고 있으며,[21] 해당 문제에 향상된 성능을 보여줄 것으로 기대된다.

Table 2는 Table 1에 해당하는 전체 데이터의 통계적 성능을 평가하기 위해 전체 대역과 각 Case마다 표적이 존재하는 대역의 선형 스펙트로그램 주파수

Table 2. Overall averaged MAE between original and synthesized active sonar signals.

	Target type	Total band (dB)	Target band (dB)
Case 1	CW	5.69	4.32
	LFM	6.13	6.60
Case 2	CW	0.93	0.20
	LFM	2.86	5.53
Case 3	CW		
	LFM	6.45	6.84
Case 4	CW	-3.20	-4.20
	LFM	0.43	1.33

Table 3. PEAQ score between original and synthesized active sonar signals.

	type	Score
Case 1	CW	-1.309
	LFM	-2.052
Case 2	CW	-2.029
	LFM	-2.105
Case 3	CW	
	LFM	-1.721
Case 4	CW	-2.103
	LFM	-2.128

계수들을 평균 MAE를 dB 크기로 평가하였다. 표에서 확인할 수 있듯이 모든 Case와 표적 종류에 대해 최대 7 dB 이내의 비교적 적은 오차로 능동 소나 신호를 모델링 하는 것을 확인할 수 있다. 하지만, CW 신호에 대해서는 표적 대역이 전체 대역보다 적은 오차를 보이고 있는 반면 LFM 신호의 경우 전체 대역이 표적 대역보다 적은 오차를 보이는 등 전체 신호에 걸쳐 공통적인 경향을 보이지 않고 불규칙적으로 분포되어 있다. 이는 앞서 언급한 표적 신호의 에너지를 과하게 부각시키는 문제와 더불어 배경 잡음 역시 실제 신호와 어느 정도 차이를 보이고 있는 것을 의미한다. 해당 문제의 해결을 위해 역시 앞서 언급한 적대적 생성 학습법을 활용하여 보다 사실적인 배경 잡음 및 표적 신호의 합성을 필요로 한다.

Table 3은 제안하는 모델의 능동 소나 신호 합성 성능의 정량적 평가를 위해 측정한 PEAQ 점수를 보여준다. 평가는 원본 소나 신호를 참조 신호, 합성된 신호를 시험 신호로 설정하여 진행하였다. 해양환경의

종류를 의미하는 네 종류의 파라미터를 따라 실험 케이스를 표와 같이 설정하였으며, 각 Case 마다 LFM, CW 펄스 데이터 100개씩 총 500개의 시험 데이터를 설정하여 그 평균값을 최종 결과값으로 설정하였다. 실제 능동 소나 신호와 합성된 신호 사이의 인지적 유사도를 (0,-4) 범위의 실수값으로 측정하게 된다. 점수가 0에 가까울수록 참조 신호와의 인지적 차이가 구분하기 어려움을 의미하며, -4에 가까울수록 그 차이가 심해짐을 말한다. 그 결과 전체적으로 -2.3점 이내의 점수로 약한 수준의 인지적 차이를 보여준다.[22] 또한, 신호대 잡음비가 강한 Case 1과 3의 경우 평균 -1점대의 청취하였을 때 거슬리지 않는 정도의 성능을 보여준 반면, 잡음 에너지가 비교적 강한 Case 2와 4의 경우는 평균 -2점대의 비교적 큰 차이를 보여주었다. 이는 본 논문에서 제안하는 능동 소나 신호 합성 모델이 인지적으로 양호한 표적 모델링 성능을 입증하지만, 이와 반대로 배경 잡음 합성에는 원본과 다소 차이가 있다는 것 역시 말하고 있다. 따라서, 앞서 설명한 바와 같이 보다 자연스러운 잡음 에너지의 합성을 위해 생성 모델을 적용하게 된다면 보다 안정적인 성능을 보장할 수 있을 것으로 기대된다.

V. 결 론

본 논문에서는 심층 신경망 기반의 능동 소나의 반사 신호를 모의 및 생성하는 모델을 제안하였다. 입력으로는 해양 및 실험 환경 파라미터를 수치 벡터로서 제공하며 출력으로는 입력 파라미터값들에 대응하는 표적 반사 신호의 선형 스펙트로그램 파형 데이터를 생성한다. 음성 합성 모델 기반의 신경망 모델을 수중 능동 소나 신호에 적합하도록 구조를 변경하는 과정을 거쳤다. 입력 벡터를 환경을 모의하여 신호를 생성하는데에 필요한 정보 벡터로 변화하여 주는 인코더, 전달 받은 정보 벡터를 바탕으로 출력을 순차적으로 생성하는 디코더, 디코딩 시에 개별 시점마다 필요한 정보만을 추출 및 가공해 제공하는 주의도 모듈까지 하여 총 3가지 서브모델로 구성되었다. 제안하는 모델은 전체 대역의 스펙트로그램과 표적 신호의 중심 에너지를 비교해본 결과 훈련된 데이터셋과 상당히 유사한 신호를 생성해내는 것이 확인되었다. 나아가 WaveNet Vocoder와 결합된 형태의 Tacotron2와 NVIDIA사의 Flowtron과 같이 성능이 개선된 형태의 가변 신호 생성 모델을 사용하여 연구된다면 더욱 복합적인 파라미터를 가지는 환경 역시 반영이 가능할 것이다.

감사의 글

이 연구는 국방과학연구소의 지원을 받아 수행된 연구임(UD210005DD).

References

1. W. C. Knight, R. G. Pridham, and S. M. Kay, “Digital signal processing for sonar,” Proc. IEEE, **69**, 1451-1506 (1981).
2. A. A. Winder, “Sonar system technology,” ITSU, **22**, 291-332 (1975).
3. H. Yang, S.-H. Byun, K. Lee, Y. Choo, and K. Kim, “Underwater acoustic research trends with machine learning: active sonar applications,” J. Ocean Eng. Technol. **34**, 277-284 (2020).
4. D. Grimmett and S. Coraluppi, “Contact-level multi-static sonar data simulator for tracker performance assessment,” Proc. the 9th International Conf. Information Fusion, 1-7 (2006).
5. B. La Cour, C. Collins, and J. Landry, “Multi-everything sonar simulator (MESS),” Proc. 9th International Conf. Information Fusion, 1-6 (2006).
6. P. A.m. De Theije and H. Groen, “Multistatic sonar simulations with SIMONA,” Proc. 9th International Conference on Information Fusion, 1-6 (2006).
7. K. D. LePage, C. H. Harrison, C. Strode, and M. Oddone, “Real-time reverberation time series simulation for embedded simulation of bistatic active sonar,” Proc. OCEANS, 1-4 (2021).
8. A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “WaveNet: a generative model for raw audio,” arXiv preprint arXiv:1609.03499 (2016).
9. S. O. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, S. Sengupta, and M. Shoeybi, “Deep voice: real-time neural text-to-speech,” arXiv:1702.07825, 1-10 (2017).
10. S. Mehta, E. Szekely, J. Beskow, and G. E. Henter, “Neural HMMs are all you need (for high-quality

attention-free TTS),” Proc. ICASSP, arXiv:2108.13320, 1-5 (2022).
11. J. Sotelo, S. Mehri, K. Kumar, J. F. Santos, K. Kastner, A. Courville, and Y. Bengio, “Char2Wav: end-to-end speech synthesis,” Proc. ICLR workshop submission, 1-6, 2017.
12. Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. Le, Y. Agiomyrgiannakis, R. Clark, and R. A. Saurous, “Tacotron: towards end-to-end speech synthesis,” Proc. Interspeech, 4006-4010 (2017).
13. B. I. Kim, H. U. Lee, and M. H. Park, “A study on highlight distribution for underwater simulated target,” Proc. IEEE Int. Symp. Ind. Electron. 1988-1992 (2001).
14. Y. Kim, J. Kim, J. Hong, and J. Seok, “The tacotron-based signal synthesis method for active sonar,” Sensors, **23**, 28 (2023).
15. V. A. Kozhevnikov and E. S. Pankratova, “Research of the text data vectorization and classification algorithms of machine learning,” ISJ Theor. Appl. Sci. **5**, 574-585 (2020).
16. W. A. Qader, M. M. Ameen, and B. I. Ahmed, “An overview of bag of words; importance, implementation, applications, and challenges,” Proc. 5th Int. Eng. Conf. IEC, 200-204 (2019).
17. B. Das and S. Chakraborty, “An improved text sentiment classification model using TF-IDF and next word negation,” ArXiv: 1806.06407 (2018).
18. C. Wang, P. Nulty, and D. Lillis, “A comparative study on word embeddings in deep learning for text classification,” Proc. 4th Int. Conf. NLPI, 37-46 (2020).
19. S. Li, S. Villette, P. Ramadas, and D. Sinder, “Speech bandwidth extension using generative adversarial networks,” Proc. ICASSP, 5029-5033 (2018).
20. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio “Generative adversarial nets,” Proc. NIPS, 2672-2680 (2014).
21. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” Proc. IEEE Conf. CVPR, 105-114 (2017).
22. A. F. Khalifeh, A. Al-Tamimi, and K. A. Darabkh, “Perceptual evaluation of audio quality under lossy networks,” Proc. International Conf. WiSPNET, 939-943 (2017).

저자 약력

▸ 김 윤 수 (Yunsu Kim)

2022년 : 창원대학교 정보통신공학부 석사
2022년 ~ 현재 : 창원대학교 정보통신공학부 박사과정

▸ 김 주 호 (Juho Kim)

2010년 : 제주대학교 해양시스템공학과 학사
2012년 : 제주대학교 해양시스템공학과 석사
2016년 : 제주대학교 해양시스템공학과 박사
2016년 ~ 현재 : 국방과학연구소 선임연구원

▸ 석 종 원 (Jongwon Seok)

1995년 : 경북대학교 전자공학과 석사
1999년 : 경북대학교 전자공학과 박사
1999년 : 한국전자통신연구원 선임연구원
2004년 ~ 현재 : 창원대학교 정보통신공학과 교수

▸ 홍 정 표 (Jungpyo Hong)

2006년 : 한국과학기술원 정보통신공학과 학사
2016년 : 한국과학기술원 전기및전자공학부 박사
2016년~2020년 : 국방과학연구소 선임연구원
2020년 ~ 현재 : 창원대학교 정보통신공학과 조교수