생체모방 수중통신을 위한 VAE 트랜스포머 기반 고래 휘슬음 생성 방법

박근호, 안종민, 김완진, 김인수, 김형문, 이상국, 이동훈

국방과학연구소

ghpark57935@add.re.kr, leedhun@add.re.kr

VAE Transformer-based Cetacean Whistle Generation Method for Biomimetic Underwater Communication

Park Geun-Ho, Ahn Jongmin, Kim Wanjin, Kim In-Soo, Lee Donghun

Agency for Defense Development

요 약

본 논문은 생체모방 수중통신을 위한 흑범고래 휘슬음 생성 모델의 구현 결과를 제시한다. 휘슬음은 다양한 특징을 보유한 비선형 주파수 변조 신호로, 특정 확률 분포를 갖는다. 딥러닝 기반 휘슬음 생성 모델은 휘슬음 데이터의 확률 분포를 따르도록 매개변수를 학습할 수 있으며, 이 모델이 효과적으로 학습되면 생체모방 수중통신 목적의 신호 송신부에 적용되거나, 데이터 증강에도 활용될 수 있다. 본 논문은 흑범고래 휘슬음의 시간-주파수 데이터에 VAE 트랜스포머 인코더-디코더를 학습한 후, 잠재 공간을 혼합 가우시안 밀도로 모델링하는 방법을 제시한다. 제시한 방법은 혼합 가우시안 밀도로 모델링한 잠재 공간에서 랜덤 변수를 샘플링하고, 이를 디코더에 입력하면 조건부 휘슬 데이터 생성이 가능하다.

Ⅰ. 서 론

고래는 의사소통 및 먹이탐색 등의 목적으로 음향 신호를 수중에서 발생시킨다. 휘슬음은 고래의 대표적인 발성 중 하나로, 다양한 구조의 시간-주파수 변화를 통해 수중 환경에서 의사소통한다[1].

휘슬음은 많은 분야에서 관심있는 연구 주제이다. 고래와 돌고래의 발성과 의사소통의 연관성을 연구하거나 종에 따른 음향 신호의 차이를 분석하기 위해 휘슬음이 활용되기도 한다[1]. 휘슬음의 시간-주파수 변화를 모방하여 송신 정보를 변조하는 생체모방 통신에서도 휘슬음은 주요 모방 대상이다[2]. 생체모방 수중통신은 통신 신호를 생체 신호로 오인하게 하여 은밀성을 보장하므로, 군사적 목적의 활용도가 높다.

본 논문에서는 생체모방 수중통신을 위해 딥러닝 모델을 통해 휘슬음을 생성하는 방안에 대해 고려한다. 휘슬음의 시간-주파수 데이터는 특정 확률 분포에서 표본화된 것으로 이해할 수 있다. 이에 따라 신경망의 매개변수를 최적화하여 휘슬음의 확률 분포를 따르도록 학습할 수 있다. 이러한 휘슬음 생성 모델은 생체모방 수중통신의 신호 송신부에 적용할 수 있으며, 데이터 증강 목적으로도 활용 가능하다.

휘슬음의 시간-주파수 벡터 데이터는 각 시간 프레임의 주파수 토큰으로 변환하면 Variational Autoencoder (VAE) 트랜스포머가 다룰 수 있다. VAE의 인코더는 입력을 압축된 잠재 변수 공간에 맵핑하며, 잠재 벡터는 다시 VAE의 디코더를 통해 원래의 시간-주파수 벡터로 복원된다.

본 논문에서는 VAE 트랜스포머의 잠재 공간 표현을 학습하고, 이 잠재 공간을 다시 혼합 가우시안 밀도로 모델링함으로써 휘슬음의 조건부 생성 가능성을 제시한다. 혼합 가우시안 밀도는 다수의 가우시안 컴포넌트로 구성되며, 각각의 가우시안 컴포넌트가 휘슬음의 특정 형태를 대표한다. 특정 컴포넌트에서 잠재 벡터를 표본화하는 방법을 통해, 휘슬음을 조건부로 생성할 수 있다.

Ⅱ. 흑범고래 휘슬음 데이터

본 논문에서는 미국 해양대기청[3]의 흑범고래 휘슬음 데이터를 활용하였다. 흑범고래 휘슬음을 원신호 스펙트로그램으로부터 추출하기 위해, [4]에서 제시한 방법을 적용하였다. 이 방법으로 총 232,177개의 개별 휘슬음을 추출했다. 그림 1에 흑범고래 휘슬음 추출 예시를 나타내었다.

추출된 휘슬 데이터의 주파수는 트랜스포머에 입력할 수 있도록 토큰화하였다. 휘슬 데이터는 2 kHz에서 12 kHz의 주파수 범위를 256개의 주파수 이산 토큰으로 나누었다. 트랜스포머가 휘슬 구간의 시작과 종료 지점을 파악할 수 있도록 Start-of-Sentence (SOS)와 End-of-Sentence (EOS)를 시작과 끝에 배치했다. SOS과 EOS는 각각 0번과 257번의 토큰을 할당하였다.

Ⅲ. 휘슬 데이터의 잠재 변수 모델링

휘슬 데이터는 VAE 트랜스포머를 통해 잠재 변수로 표현될 수 있다. 잠재 변수로 변환된 데이터는 다시 혼합 가우시안 분포 모델을 통해 명시적인 확률 분포로 모델링 가능하다. 본 장에서는 VAE 트랜스포머 학습 방법과 혼합 가우시안 분포 모델에 대해 각각 다룬다.

(1) VAE 트랜스포머 학습

먼저 신경망 구조는 일반적인 인코더-디코더의 트랜스포머 블록 구조로 구성했다. 인코더는 휘슬 데이터 입력을 받아 잠재 변수로 표현하는 방법을 학습한다. 인코더는 SOS 토큰을 포함한 최대 512개 길이의 휘슬 데이터를 입력받아 4개의 트랜스포머 블록으로 512×64차원 임베딩 벡터를 출력한다. 이 중에서 첫 번째 임베딩 벡터를 Fully Connected Layer (FCN),

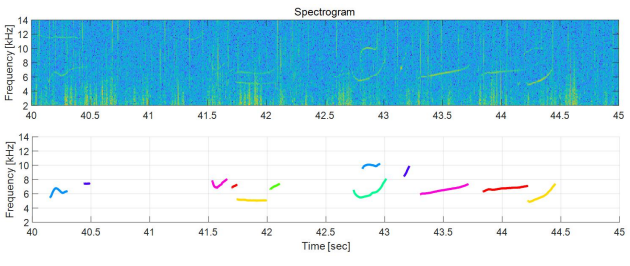


그림 1. 흑범고래 휘슬음 스펙트로그램 및 추출 예시.

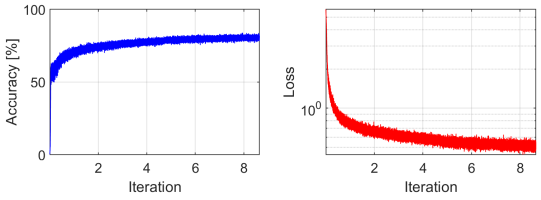


그림 2. VAE 트랜스포머의 손실 및 정확도 학습 곡선

swish layer, FCN으로 처리하고, 재매개변수화를 통해 잠재 변수를 샘플링하도록 했다.

디코더는 자기 회귀 방식으로 휘슬 데이터를 생성하기 위해, self-attention의 인과적 처리 방식을 적용하도록 구성했고, 잠재 변수를 각 트랜스포머 블록으로 입력받게 된다. 디코더는 인코더와 동일한 입력을 받아 4개의 트랜스포머 블록을 통과시켜 257차원의 확률 질량을 출력한다. 디코더의 트랜스포머 블록에서 self-attention 계층은 causal mask를 적용하였다. 길이가 다른 데이터를 배치 단위로 학습하기 위해, 0 값을 패딩하였으며, 이에 따라 0 패딩에 대한 mask를 입력하여 처리했다.

전체 손실함수는 Cross-entropy loss와 Kullback-Leibler (KL) 손실의 가중합으로 구성하였다. KL 손실의 가중치 β는 10^{-4}으로 설정했다.

그림 2는 VAE 트랜스포머의 손실 및 정확도 학습 곡선을 나타낸 것으로, 정확도와 손실이 모두 일정하게 수렴하는 것을 볼 수 있다. 학습을 위해, Epoch는 50, minibatch 크기는 128, 학습률은 10^{-4}로 설정했으며, Adam optimizer를 적용했다.

전체 데이터의 95%는 학습, 5%는 테스트에 활용하였다. 테스트 결과 정확도는 80.27%, 손실은 0.5213으로, 훈련 결과와 비슷하게 나타나 과적합은 나타나지 않았다고 판단했다.

(2) 혼합 가우시안 분포 모델링

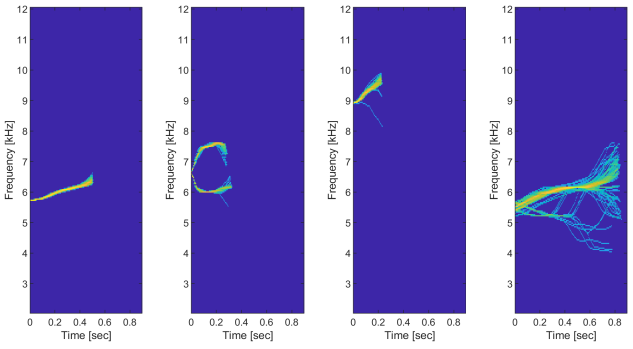
휘슬 시간-주파수 데이터는 VAE 트랜스포머를 통해 64차원의 잠재 변수로 변환하였다. 총 232,177개의 잠재 변수와 Expectation Maximization (EM) 알고리즘을 통해 혼합 가우시안 분포의 매개변수를 추정하였으며, 가우시안 컴포넌트의 수는 12,000개로 설정하였다.

Ⅳ. 휘슬 데이터의 조건부 생성

휘슬 데이터의 잠재 공간을 12,000개의 가우시안 분포로 모델링하였으므로, 잠재 공간에서의 잠재 변수 표본화 방법에 따라 무조건부와 조건부 생성이 가능해진다. 무조건부 생성은 12,000개의 가우시안 분포 중 하나를 mixing coefficient에 따라 샘플링한 후, 해당 가우시안 분포에서 잠재 변수를 순차적으로 샘플링한다. 반면 조건부 생성은 특정 가우시안 분포를 선택한 후, 무조건부와 동일한 과정을 수행할 수 있다.

그림 3은 k번째 가우시안 컴포넌트의 잠재 변수를 통해 100개의 휘슬을 생성하여 누적한 결과를 나타낸 것으로, 특정 컴포넌트의 잠재 변수가 일정한 형태의 휘슬음을 생성하도록 유도하는 것을 볼 수 있다. 그림 3의 유도 결과에서 Temperature는 0.4로 설정했다.

조건부 생성 결과 (그림 3)에서 확인할 수 있는 특징은 하나의 컴포넌트에서 단일의 휘슬 형태만 표본화되지 않을 수 있다는 점이다. 그림 3의 (b)와 (d)는 이러한 현상을 잘 나타내고 있다. 이 결과는 두 가지 가능성을 시사하는데, 하나는 가우시안 컴포넌트의 수가 부족해서 다른 휘슬 형태의 두 클러스터를 하나의 가우시안 분포로 모델링했을 가능성이다. 이 경우 더 많은 수의 가우시안 분포를 활용해야 하지만, 이미 많은 수의 가우시안 컴포넌트 수인 12,000개를 적용하였으므로, 혼합 가우시안 분포가 아닌 다른 잠재 변수 모델링 방안이 더 적합할 수 있다. 두 번째는 잠재 변수 중 하나가 시작 주파수를 축으로 하는 대칭을 의미하는 경우이다. 그림 3



(a) (b) (c) (d)

그림 3. 휘슬 조건부 생성 결과: (a) k=1, (b) k=4,329, (c) k=6,839, (d) k=10,004.

의 (b)는 시작 주파수를 축으로 시간-주파수 패턴이 두 그룹으로 나뉜다. 반면 (d)는 조금 더 복잡한데, 시작 주파수 부근에서 한 번 분기하며, 0.6초 부근에서 다시 한번 나뉜다. 이것이 VAE 트랜스포머 기반의 잠재 변수를 혼합 가우시안 분포로 모델링할 때의 특징일 가능성도 배제할 수 없다.

Ⅴ. 결론

본 논문은 생체모방 통신을 위한 VAE 트랜스포머 기반의 고래류 휘슬의 조건부 생성 결과를 제시하였다. 그리고 잠재 변수 공간을 혼합 가우시안 밀도로 모델링하고 각 가우시안 컴포넌트를 제어하여 특정 형태 생성이 가능한 것을 확인하였다. 추후에는 해당 트랜스포머 기반의 휘슬 생성 방법의 특징에 대해 더 상세히 파악할 예정이며, 트랜스포머의 자기 회귀 특성에 따라 연산량이 발생하는 문제를 개선하기 위한 방안도 연구할 계획이다.

ACKNOWLEDGMENT

이 논문은 2026년 정부(방위사업청)의 재원으로 국방과학연구소가 수행한 미래도전국방기술 연구개발사업의 연구 성과임(No.915084201).

참 고 문 헌

[1] Y. G. Yoon, *et. al.*, “Study of Acoustic Characteristics of Common Dolphins *Delphinus delphis* in the East Sea,” *KFAS*, vol. 50, no. 4, pp. 406-412, 2017.

[2] S. Seol, *et. al.*, “Research trends of biomimetic covert underwater acoustic communication,” *The Journal of the Acoust. Soc. of Korea*, vol. 41, no. 2, pp. 227-234, 2022.

[3] NOAA Pacific Islands Fisheries Science Center. 2022. Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS) towed array data. Edited and annotated for the 9th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals Using Passive Acoustics (DCLDE 2022). NOAA National Centers for Environmental Information. https://doi.org/10.25921/e12p-gj65(https://doi.org/10.25921/e12p-gj65), accessed: Nov. 30, 2023

[4] G.-H. Park, J. Ahn, W. Kim, I. S. Kim, and D. H. Lee, “Implementation of a neural network for cetacean whistle frequency detection based on data synthesis,” in *Proc. 2024 Winter Conf. Korean Inst. Commun. Inf. Sci. (KICS)*, Gangwon, South Korea, Feb. 2025, pp. 690-691.