이중 신경망을 활용한 군 무전 통신체계의 잡음내성 향상 방안에 대한 연구

박성진*
*LIG넥스원
(seongjin.park@lignex1.com)

요 약

전시 상황에서 무전 통신 시 비행기, 포탄, 총 소리 등 여러 가지 소음이 첨가된 음성이 송신될 수 있다. 또한 아날로그 음성 신호를 전송하는 과정에서 외부 잡음이 유입되어 노이즈가 이중으로 유입된 신호가 최종 수신된다. 이와 같은 과정으로 인해 송수신부 간 통신 무결성이 저하될 수 있다. 이를 개선하고자 송신 무전기와 수신 무전기에 Audio Denoising 신경망을 활용하여 통신 잡음을 개선할 수 있는 방법을 검토 및 검증한다.

1. 서 론

군 무선 통신체계에서 가장 우선시 되는 것은 정확한 정보 전달 능력이다. 이를 위해선 신호 전송 중 유입되는 왜란에 대한 내성도 중요하지만 전송하고자 하는 정보자체의 품질이 우수해야 한다. 전시 시엔 총, 포탄, 전투기, 전차 소리 등 다양한 노이즈가 존재하기에 깨끗한 음성신호를 무전기로 입력하기란 쉽지 않다. 또한 깨끗한 음성이 입력되어도 전송 중 외부에서 유입되는 자연잡음, 인공잡음의 영향으로 신호가 혼탁해진다. 이를 개선하고자 노이즈가 유입된 음성신호를 필터링할 수 있는 Audio Denoising 신경망의 활용 방안을 제시한다.

2. 본론

2.1 데이터 전처리

사람이 발성할 수 있는 음성의 대역폭은300Hz~4kHz로 가청대역폭이 20Hz ~ 20kHz이상인 점을 감안하면 낮은 대역폭을 갖는다. 음성 데이터에서 사람의 목소리만 분석하기 위해선 데이터를 시간영역에서 주파수영역으로 변환해야 한다. 주파수 영역으로 변환하는 정형적인 방식은 전체 시간영역에 대해 FT(Fourier Transform)을 취하는 것이다. 하지만 FT의 결과는 어느 시점에 어떤 주파수 성분이 존재하는지 알 수 없기 때문에 음성의 연속성이란 정보가 손실된다. 그래서 시간을 짧게 나누어 변환하는 STFT(Short Time Fourier Transform)로 전처리한 결과를 입력 데이터로 사용하였다. STFT는 짧은 구간으로 시간을 나눌 때 Overlap 영역이 존재하여 음성구간 간 연속성을 높이며 변화하는 주파수 경향을 파악할 수 있다.

학습에 필요한 노이즈가 없는 음성 데이터는 [6]에서 제공하는 한국 남녀의 음성데이터를 사용했다. 해당 데이터에 전시 시 발생되는 소음과 신호 전송 시 유입되는 외부잡음을 각각 합성하여 노이즈가 있는 데이터로 생성했다.

수신부에서 필터링해야 되는 노이즈는 음성 입력과 동시에 함께 유입되는 소음이다. 전시 상황에서 발생할 수 있는 25종의 효과음을 랜덤하게 배열 후 음성 데이터에 합성해 수신부에서 필터링해야 하는 노이즈 음성 데이터 셋을 추출했다. 수신부로 부터 전송된 신호에는 외부의 다양한 왜란으로 인해 발생하는 백색소음이 유입된다. 이와 같은 환경을 유사하게 재현하고자 정규분포를 따르는 노이즈를 생성한 후 음성 신호와 합성시켜 송신부에서 필터링해야 하는 학습 데이터 셋을 추출하였다.

2.2 학습 신경망 아키텍처

지도학습은 입력 데이터를 목표 데이터에 근사하게 출력시키는 학습 방식이다. 음성데이터를 지도 학습하기 위해선 노이즈가 없는 음성 데이터가 필요하지만 이를 실제로 채취하기란 쉽지 않다. 음성을 측정할 때 기계의 잡음 혹은 주변 환경의 소음으로 노이즈가 조금 낀 데이터가 수집될 가능성이 높기 때문이다. 노이즈가 조금 낀 데이터를 목표로 지도학습하면 노이즈가 수반된 데이터가 출력되도록 학습될 수 있다. 이러한 문제점을 보완 하고자 [4]에서 언급되는 Noise2Noise 기반의 신경망을 사용한다. 해당 신경망은 입력 데이터인 Nosiy Input이 Noisy Target으로부터 생성된 Average gradient 사용해 Clean Target을 유추하는 비지도 학습이다. 고로 원본 데이터에 숨어있는 노이즈가 존재 하더라도 노이즈가 존재하는 두 이미지의 평균을 최종 출력으로 학습하므로 숨어있는 노이즈에 대한 영향력이 감소된다.

Noise2Noise 신경망은 이미지 데이터를 Denoising 하도록 설계된 신경망임으로 음성데이터에서 추출된 복소수 구성의 STFT를 적용하기엔 적합하지 않다. STFT는 이미지의 RGB값 대신 주파수의 진폭과 위상으로 데이터가 구성된다. 진폭 값을 연산할 때 마이너스를 곱하면 위상이 반전되는 것과 같은 의미이다. 기존 MSE 기반의 LossFuction을 사용하면 진폭에 제곱을 취하기 때문에 진폭에 마이너스가 정보가 사라져 위상이 반전된 신호도 같은 오차 패널티를 받게 된다. 이러한 문제점을 개선하고자 [5]에서는 Phase Estimation을 좀 더 효과적으로 할 수 있는 Weighted SDR Loss Fuction을 사용하여 Complex-valued Masking에 우수한 모델임을 설명하고 있다. 현재 학습 데이터에 적합한 아키텍처이므로 해당 모델을 사용하여 노이즈 필터링 환경을 구성하였다.

2.3 학습 및 활용 방안

무전기의 송신부에선 음성과 함께 들어오는 총, 포탄, 전차, 전투기 소음을 필터링한 후 송신시켜야 한다. 송신된 신호는 전송 중 여러 왜란의 영향으로 백색소음이 유입되며 수신 무전기에선 이를 필터링해야 한다. 필터링 운영의 성능 검증을 다음 3 가지 방법으로 실험해 보았다.

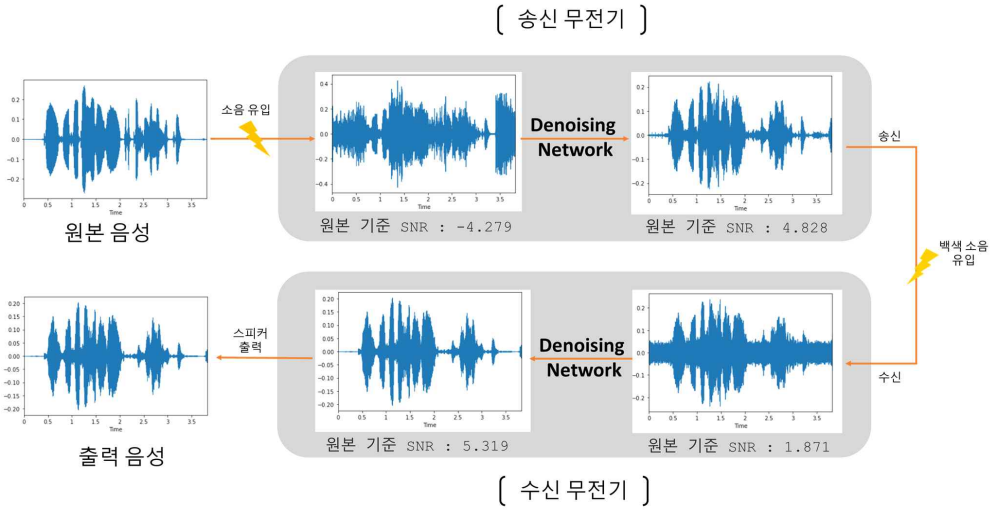

그림 1 "방법 3"의 학습법을 사용한 테스트 데이터의 노이즈 필터링 운영 계통도 및 출력 결과

방법 1 – 송신부, 수신부에 동일하게 학습된 신경망과 (원본 음성, 원본 음성+음성 입력단 소음) or (원본 음성, 원본 음성 + 백색 소음) 로 구성된 데이터 셋을 사용했다.

방법 2 - 수신부와 송신부 신경망을 각각 할당하여 이중으로 학습하는 방법이다. 데이터는 각각 송신부 신경망은 (원본 음성, 원본 음성+음성 입력단 소음) 데이터 셋을, 수신부 신경망은 (원본 음성, 원본 음성 + 백색 소음) 데이터 셋을 사용했다.

방법 3 - "방법 2"와 동일하게 이중 신경망을 사용하고 수신부 학습 데이터 셋도 동일하다. 차이가 있는 부분은 송신부 신경망을 학습할 때 (원본 음성, 송신부의 필터링된 신호+백색잡음)으로 구성된 데이터 셋을 사용하여 백색 소음을 필터링함과 동시에 수신부에서 필터링된 신호를 한 번 더 필터링 시킨다.

2.4 학습 결과 및 비교 분석

최종학습 결과, 2.3에서 언급된 세 번째 운영 방법이 SNR 수치적으로 가장 우수했다. 학습 방법에 따른 SNR을 비교하면 표 1과 같다.

방법	신경망 종류	입출력	수신부 학습 데이터 셋	송신부 학습 데이터 셋	SNR
1	단일	출력	원본 음성	수신부와 동일	4.05
		입력	원본 음성 +소음 +백색소임		
2	이중	출력	원보 음성	원본 음성	4.22
		입력	원본 음성 +소음	원본 음성 +백색소음	
3	이중	출력	원본 음성	원본 음성	4.48
		입력	원본 음성 +소음	수신부 필터링 신호 +백색소음	

표 1 학습 및 신경망 활용 방식에 따른 최종 SNR 비교

"방법 3" 사용하여 학습된 신경망으로 테스트 음성 신호를 필터링 했을 때와 신경망을 사용하지 않고 노이즈가 유입된 최종 출력 신호를 비교하면 "그림 2"와 같다. 필터링을 사용하지 않아 노이즈가 유입된 출력 신호는 SNR이 –6.807이고 Audio Denising 신경망을 사용한 환경의 출력 신호 SNR은 5.319로 큰 차이를 보인다.

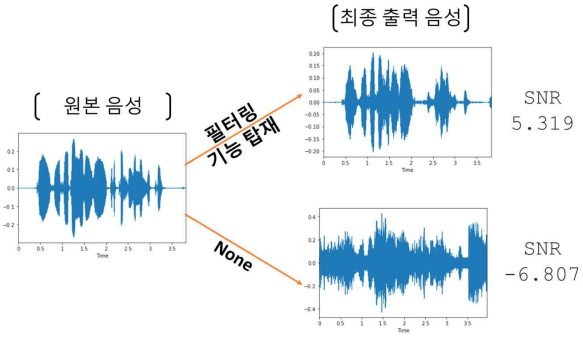

그림 2 "방법 3" 기준 신경망 유무에 따른 출력 결과

3. 결론

수신, 송신 필터링 동작에 특화 될 수 있도록 분리 학습 시킨 이중 신경망을 사용하면 노이즈 유입에 내성을 가진 통신체계를 구축할 수 있다. 전시 시 신경망 필터링 기능이 탑재된 무전기를 사용하여 통신 무결성을 향상 시켜 전투력 증강에 기여할 수 있다.

참 고 문 헌

[1] 김동욱, 문행규, 『전파통신』, 사이버북스(2015. 9. 2)

[2] "주파수 스펙트럼", 『중안전파관리소』, https://www.crms.go.kr/lay1/S1T1010C1015/contents.do

[3] "STFT(Short Time Fourier Transform)", 『naverblog』, 2017. 2. 15, https://m.blog.naver.com/vmv-tech/220936084562

[4] Lehtinen, Jaakko, et al. "Noise2Noise: Learning image restoration without clean data."

[5] Kashyap, Madhav Mahesh, et al. "Speech Denoising without Clean Training Data: a Noise2Noise Approach."

[6] "자유대화 음성(일반남녀)", 『AI Hub』, 2022. 05 https://www.aihub.or.kr/aihubdata/data