https://doi.org/10.33778/kcsa.2024.24.2.201

군사적 환경에서 음성인식 모델의 취약성에 관한 연구★

원 엘 림*, 나 성 중**, 고 영 진***

요 약

목소리는 인간의 의사소통에서 중요한 요소로, 음성인식 모델의 발전은 인공지능의 중요한 성과 중 하나이며 최근 인간의 생활에 다방면으로 사용되고 있다. 음성인식 모델의 활용은 군사분야에서도 피해갈 수 없는 과제이다. 하지만 인공지능 모델의 군사적 활용 이전에 모델의 취약성에 대한 연구가 필요하다. 본 연구에서는 다국적 음성인식 모델인 Whisper의 군사적 활용 가능성을 알아보기 위해, 전장소음, 잡음, 적대적 공격에 대한 취약성을 평가하였다. 전장소음을 포함하는 실험에서는 Whisper의 성능 저하가 크게 나타났으며, 평균 72.4%의 문자 오류율(CER)을 기록하여 군사적 활용에 어려움이 있는 것으로 나타났다. 또한, 잡음을 포함하는 실험에서는 낮은 강도의 잡음에 대해 Whisper가 강건하였으나, 높은 강도의 잡음에서는 성능이 저하되었고, 적대적 공격 실험에서는 특정 입실론 값에서 취약성이 드러났다. 따라서 Whisper 모델을 군사적 환경에서 사용하기 위해서는 파인튜닝, 적대적 훈련 등을 통해 개선이 필요하다는 것을 시사한다.

Study on the Vulnerabilities of Automatic Speech Recognition Models in Military Environments

Elim Won*, Seongjung Na**, Youngjin Ko***

ABSTRACT

Voice is a critical element of human communication, and the development of speech recognition models is one of the significant achievements in artificial intelligence, which has recently been applied in various aspects of human life. The application of speech recognition models in the military field is also inevitable. However, before artificial intelligence models can be applied in the military, it is necessary to research their vulnerabilities. In this study, we evaluates the military applicability of the multilingual speech recognition model "Whisper" by examining its vulnerabilities to battlefield noise, white noise, and adversarial attacks. In experiments involving battlefield noise, Whisper showed significant performance degradation with an average Character Error Rate (CER) of 72.4%, indicating difficulties in military applications. In experiments with white noise, Whisper was robust to low-intensity noise but showed performance degradation under high-intensity noise. Adversarial attack experiments revealed vulnerabilities at specific epsilon values. Therefore, the Whisper model requires improvements through fine-tuning, adversarial training, and other methods.

Key words : Automatic Speech Recognition, Battlefield noise, Adversarial attack, Vulnerability

접수일(2024년 05월 20일), 수정일(1차: 2024년 05월 31일), 게재확정일(2024년 06월 07일)

★ 본 논문은 육군사관학교 화랑대연구소의 2024년도 논문게재지원비 지원을 받아 연구되었음.

* 육군 지능정보기술단 체계관리SW분석장교(주저자)

** 국방대학교 군사운영분석 박사과정(공동저자)

*** 육군사관학교 컴퓨터과학과 강사(교신저자)

1. 서 론

목소리는 인간의 의사소통에서 가장 중요한 요소 중 하나이다. 문자를 사용한 의사소통보다는 즉각적이며, 몸짓을 통한 의사소통보다는 훨씬 정확하다. 이러한 특성으로 인간은 음성을 통한 상호작용을 더 선호하며, 인공지능의 발전에 발맞추어 뛰어난 성능을 나타내는 자동 음성 인식(Automatic Speech Recognition, ASR)모델이 나타나게 되었다[1].

1950년대의 초기 음성인식 기술은 단순한 숫자나 단어 인식에 국한되었지만, 1980년 이후 은닉 마르코프 모델 기반의 음성인식 모델이 제시되었고, 2017년 이후에는 Transformer 기반의 음성인식 모델이 나타나게 되면서 복잡한 문장과 다양한 언어를 인식할 수 있는 모델로 발전해 왔다. 오늘날, 음성인식 모델은 스마트폰, 네비게이션 등 우리의 일상 속에 확산되어 우리의 생활을 더욱 편리하고 효율적으로 만들어 주고 있으며, 인간과 기계 간의 상호작용 방식을 혁신적으로 변화시키고 있다.

육군에서도 Army tiger 4.0 정책에 따라 워리어 플랫폼(Warrior Platform)라고 불리는 첨단 개인전투체계를 구축하고 있으며, 전장에서 지휘관을 보좌할 수 있는 인공지능 기반의 참모시스템을 개발하고 있다. 또한, 연합작전을 위하여 다국적 음성인식 모델의 활용은 필수적이지만 음성인식 모델은 여전히 취약점을 가지고 있다. 예로, 음성인식 모델은 일반적으로 소음 또는 잡음과 함께 인식될 경우, 인식률이 저하된다[2]. 또한, 최근에는 딥러닝 모델들의 적대적 공격에 대한 취약성이 대두되면서 딥러닝을 기반으로한 음성인식 모델들의 문제점도 나타나고 있다[3, 4].

본 연구에서는 다국적 음성인식 모델을 군사적 활용 이전에 모델이 가진 취약점을 파악하고, 향후 다국적 음성인식 모델을 군사적 활용 시 개선 및 고려해야 하는 사항을 제시한다.

2장은 관련연구로 음성인식 모델의 소개와 음성인식 모델의 강건성 및 적대적 공격에 대하여 서술한다. 3장에서는 다국적 음성인식 모델의 취약성을 실험하기 위한 데이터에 대해 서술하며, 4장에서는 실험결과를 통한 다국적 음성인식모델의 취약점에 대해서 서술한다.

본 논문의 기여는 다음과 같다. 한국어 음성데이터를 사용하여 다국적 음성인식 모델인 Whisper의 모델크기(Medium, Small, Tiny)별 전장소음 및 잡음에 대한 취약성을 평가한다. 또한, 적대적 공격에 대한 취약성을 평가하여, 향후 군사적 목적으로 다국적 음성인식모델의 활용 가능 여부와 향후 연구방향을 제시한다.

2. 관련연구

2.1 음성인식모델

2010년대 음성인식 분야에서는 딥러닝 기반 접근법이 제안되기 시작하였고, 2014년에 발표한된 Deep Speech[5]는 딥러닝을 활용한 종단형(End-to-End) 음성인식 모델의 초기 사례로, 음성데이터를 직접 텍스트로 변환하는 접근 방식을 도입하여 주목받았다. 2017년에는 Transformer[6]가 발표되었고, 이후 대규모 데이터를 기반으로 음성인식 모델의 성능이 획기적으로 향상되었다. 이러한 모델들은 복잡한 전처리 과정을 요구하지 않으며, 기존의 전통적인 모델보다 훨씬 높은 성능을 보였다. 대표적인 예로는 OpenAI의 다국어 음성인식 모델인 Whisper[7]와 Facebook의 Wav2Vec 2.0[8]이 있다.

2.2 음성인식모델의 강건성

Transformer 기반의 종단형 모델은 성능향상을 위해서 모델의 크기와 파라미터의 수가 증가되는 단점이 있지만 이를 극복하기 위해, 대량의 음성데이터를 미리 학습시켜 사전학습된 모델을 사용해도 좋은 성능이 나오도록 연구되었다. 사전학습된 모델들은 학습 간 소음이 포함된 데이터도 학습하여 일반적인 소음이 존재하는 환경에서도 우수한 성능을 발휘한다. 본 연구에서 활용할 Whisper 모델 또한 일반적인 소음에 대해 강건하다고 할 수 있다[7, 9]. Whisper는 680,000시간의 다국적 음

성데이터를 활용하여 약한 지도학습(Weakly Supervised Learning)을 수행하였으며, 이러한 대규모 데이터는 다양한 환경에서 수집되어, 음성인식 모델이 소음 환경에서도 잘 작동하도록 훈련되었다. 하지만 일반적인 소음 환경이 아닌 특수한 환경에서는 성능이 저하 되는 상황이 발생한다[10].

2.3 음성인식모델에 대한 적대적 공격

딥러닝 기반 모델들의 높은 성능에도 불구하고, 최근 이러한 모델들이 적대적 예제에 취약하며 신뢰할 수 없다는 연구들이 진행되었다. 적대적 예제는 입력 데이터에 미세한 노이즈를 추가하여 인공지능 모델의 성능을 저해할 수 있는 데이터를 의미하며, 인간에 의해서는 탐지가 어렵다. 일반적으로 적대적 예제는 이미지 도메인에서 활발하게 연구가 되었지만, 최근에는 음성인식 분야에서도 활발히 진행되고 있다[3, 4].

적대적 공격(Adversarial Attack)은 일반적으로 화이트박스 공격과 블랙박스 공격으로 나눌 수 있다. 화이트박스 공격은 모델의 밝혀진 구조와 파라미터 정보를 활용하여 공격을 수행하는 반면, 블랙박스 공격은 입력과 출력 정보만을 통해서 모델을 속이는 방법을 사용한다[11]. 최근 공개된 음성인식 모델은 구조가 공개된 경우가 많아 본 연구에서는 화이트박스 공격으로 실험을 진행한다.

화이트 박스 공격의 알고리즘은 다양하지만, 그중에 가장 널리 사용되는 것이 FGSM(Fast Gradient Sign Method)이다. FGSM의 주요 강점은 단순함과 계산 효율성이다. 이 방법은 한 번의 학습으로 적대적 예제를 생성할 수 있어 매우 빠르게 수행된다. 그러나 이러한 단순함 때문에 FGSM은 비교적 약한 공격 방법으로 간주될 수 있다. FGSM의 한계를 극복하고 더욱 강력한 적대적 예제를 생성하기 위해 PGD(Projected Gradient Descent)라는 방법이 제안되었다. PGD는 FGSM와 다르게 여러 번의 작은 스텝을 통해 적대적 예제를 점진적으로 개선한다[11]. 본 연구에서 사용되는 적대적 예제의 생성에는 PGD를 활용하였으며 3.1.3에서 기술되어 있다.

3. 실험환경 및 데이터

본 연구에서는 다국적 음성인식 모델의 군사적 환경 하 활용 가능성을 테스트하기 위하여 실험환경을 세 가지로 조성하였다. 첫 번째, 전장소음에 대한 취약성을 평가하기 위하여 전장에서 발생할 수 있는 소음(총, 포탄 소리)과 한국어 음성데이터를 합성하여 실험을 진행하였다. 지상전, 해상전, 공중전 등 전쟁 양상에 따라 다양한 전장 소음이 발생할 수 있지만, 본 연구에서는 지상전 시 기본적으로 발생하는 총기 소리와 포탄 소리에 대한 취약성 실험을 진행하였다. 두 번째, 무선통신 간 발생할 수 있는 잡음에 대한 취약성에 대하여 실험하고자 가우시안 노이즈를 한국어 음성데이터와 합성하여 실험을 진행하였다. 세 번째, 적대적 공격에 대한 취약성에 대해서 실험하기 위해 PGD 기법으로 적대적 예제를 생성하여 모델의 취약성을 평가하였다. 아래 3.1에서는 실험을 위한 데이터 생성에 관하여 기술하였으며, 3.2에서는 실험에 사용되는 다국적 음성인식 모델인 Whisper에 대해 기술하였다. 또한, 3.3에서는 음성인식 모델의 인식률 및 취약성을 평가하기 위한 평가지표에 대해서 기술하였다.

3.1 음성데이터

실험에 사용한 한국어 데이터는 AI Hub에서 제공하는 한국어 강의 데이터로 2023년 12월 29일 전체 공개되었으며 인문, 공학, 예체능 등 다양한 강의 환경에서 수집된 데이터로 구성되어 있다. 본 실험에서는 해당 데이터에서 임의로 추출한 10시간 분량의 음성데이터를 사용하였으며, 적대적 예제 생성에는 임의로 추출한 1시간 분량의 음성데이터를 사용하였다.

3.1.1 전장소음이 섞인 음성데이터 합성

한국어 음성데이터에 전장소음을 추가하기 위하여 AI Hub에서 제공하는 인공적 발생 비언어적 소리 데이터를 사용하였다. 해당 데이터는 125가지의 다양한 소음데이터로 구성되어 있으며, 총소리, 포탄 소리 등 전장소음과 관련된 데이터가 포함되어 있다. 본 실험

에서는 다국적 음성인식 모델의 전장환경 내 활용 가능성과 취약성을 알아보기 위해 서로 다른 100개의 총, 포탄 소리를 임의로 선정하여 한국어 강의 데이터와 합성하였다. 이때 실제 전장환경을 가정하기 위해서 전장소음은 한국어 음성데이터의 최대 진폭의 5배로 설정하여 데이터를 합성하였다.

3.1.2 잡음이 섞인 음성데이터 합성

군에서 무선통신 간 다양한 이유로 잡음이 발생한다. 지형이나 기상조건에 따라 음성데이터의 품질이 저하되기도 하며, 송·수신기 간의 거리나 주고받는 신호의 강도에 의해 음성데이터의 품질이 달라지기도 한다. 이러한 조건에서의 음성인식 모델의 취약점을 알아보기 위해 한국어 음성데이터에 가우시안 노이즈를 추가하여 데이터를 합성하였다. 이때, 가우시안 노이즈를 잡음의 강도(Noise level)별로 구분하였으며 원본 음성데이터(x)의 최대 진폭대비 잡음의 강도를 곱하여 표준편차(σ)로 사용하였다. 잡음이 추가된 음성데이터(y)는 아래 수식(1)과 같으며, 데이터 생성에 사용한 잡음의 강도(Noise level)는 0.01, 0.05, 0.1 세 가지이다.

$$y = x + \eta,\ \eta \sim N(0, (Noise\ \ level \times A_{\max})^2) \quad (1)$$

3.1.3 적대적 예제

최근 딥러닝 기반의 모델들이 적대적 공격에 취약하다는 연구에 기반하여, 군사적으로 활용되는 인공지능 모델들의 적대적 공격에 대한 취약성 평가는 필수가 되었다. 따라서 본 연구에서도 PGD 기법을 통하여 적대적 예제를 생성하고 다국적 음성인식 모델의 취약성을 평가하고자 한다. 본 실험을 위한 적대적 예제 생성에 활용된 PGD 기법은 그림 1과 같으며, $\alpha = 0.01$, $K = 100$로 설정하고, 입실론(ϵ)의 경우에 $\epsilon = 0.01$, 0.1, 0.3 세 가지 경우로 나누어 적대적 예제를 생성하였다.

Algorithm 1 Projected Gradient Descent (PGD) Attack

Require: Model f, Original input x, True label y, Perturbation limit ϵ, Step size α, Number of iterations K
Ensure: Adversarial example x_{adv}

$\delta \leftarrow$ Random noise within $[-\epsilon, \epsilon]$
$x_{adv} \leftarrow x + \delta$
for $k = 1$ to K **do**
  Compute loss $L(x_{adv}, y)$
  Compute gradient $g \leftarrow \nabla_x L(x_{adv}, y)$
  Update $x_{adv} \leftarrow x_{adv} + \alpha \cdot \mathrm{sign}(g)$
  Project x_{adv} into the ϵ-ball of x
  Clip x_{adv} to ensure valid range
end for
return x_{adv}

(그림 1) PGD 알고리즘

<표 1> Whisper의 모델별 요구성능 및 파라미터 수

Models		Required VRAM	The number of parameters
Whisper	Large	~10 GB	1550M
	Medium	~5 GB	769M
	Small	~2 GB	244M
	Base	~1 GB	74M
	Tiny	~1 GB	39M

3.2 음성인식모델

Open AI에서 제공하는 음성인식 모델인 Whisper는 GPU 요구성능과 파라미터의 개수에 따라 Large부터 Tiny까지 구분된다. 실험에서 사용될 모델은 군사적 활용 가능성을 고려하여 Medium, Small, Tiny 세 가지 모델을 사용하였다.

3.3 평가지표

음성인식 모델의 성능평가에는 문자 오류율(Character Error Rate, CER)과 단어 오류율(Word Error Rate, WER) 두 가지가 주로 사용된다. 한국어 음성인식 평가에는 띄어쓰기 문제로 인하여 경계가 모호한 경우가 많아 WER보다는 CER을 활용하는 것이 일반적이다[9]. 적대적 공격의 경우에는 적대적 성공률(Attack Success Rate), 오분류율 등을 일반적으로 성능평가에 사용하지만, 한국어 음성인식에 대한 성능평가에는 CER을 사용하는 것이 더 적합하기 때문에 본 실험에서도 CER을 평가지표로 사용한다.

$$CER = \frac{S_c + D_c + I_c}{N_c} \quad (2)$$

수식 (2)에서 S_c(Substitution)는 잘못 대체된 글

자의 개수, D_c(Deletion)는 삭제된 글자의 개수, I_c(Insertion)는 잘못 삽입된 글자의 개수, N_c는 정답 텍스트의 총 글자의 개수를 의미한다.

4. 실험결과

아래 실험 4.1, 4.2, 4.3은 3장의 3.1.1, 3.1.2, 3.1.3의 데이터를 Whisper의 Medium, Small, Tiny 모델에 테스트한 결과이다. 먼저 원본 음성데이터에 대한 Whisper의 성능을 측정하였을 때, CER 기준 Medium 모델은 23.7%, Small 모델은 25.2%, Tiny 모델은 44.3%로 측정되었다. 해당 결과를 기준으로 전장소음, 잡음, 적대적 공격에 대한 Whisper의 성능 저하 결과를 4.1, 4.2, 4.3에서 확인할 수 있다.

4.1 전장소음에 대한 취약성

Whisper의 모델별 한국어 음성데이터에 대한 인식률은 표 2와 같이 CER 기준 Medium 모델은 23.7%, Small 모델은 25.2%, Tiny 모델은 44.3%로 나타났다. 하지만, 전장소음이 추가된 음성데이터에 대해서는 Medium 모델은 52.9%, Small 모델은 64.4%, Tiny 모델은 인식불가로 나타났다.

4.2 잡음에 대한 취약성

잡음이 추가된 음성데이터에 대한 인식률은 표 3와 같이 CER 기준 Medium 모델은 평균 36.7%, Small 모델은 42.9%, Tiny 모델은 80.2%로 나타났다. 즉, 잡음이 섞이지 않은 음성데이터를 인식할 경우와 비교하여 각각 13%, 17.7%, 35.9%가 저하되었다.

4.3 적대적 공격에 대한 취약성

잡음이 추가된 음성데이터에 대한 인식률은 표 4와 같이 CER 기준 Medium 모델은 평균 36.7%, Small 모델은 42.9%, Tiny 모델은 80.2%로 나타났다. 즉, 잡음이 섞이지 않은 음성데이터를 인식할 경우와 비교하여 각각 13%, 17.7%, 35.9%가 저하되었다. 표 5에서는 적대적 공격이 성공한 사례로, 입력된 음성과 완전히 다른 결과를 얻는 것을 확인할 수 있다.

<표 2> 전장소음이 추가된 음성데이터에 대한 인식률

Whisper	Original Speech	Original Speech + Battlefield Noise
	CER	
Medium	23.7%	52.9%
Small	25.2%	64.4%
Tiny	44.3%	N/A

<표 3> 잡음이 추가된 음성데이터에 대한 인식률

Whisper	Original Speech	Original Speech + Gaussian Noise	
	CER	Noise Level	CER
Medium	23.7%	0.01	26.0%
		0.05	34.5%
		0.1	49.7%
Small	25.2%	0.01	27.1%
		0.05	40.6%
		0.1	61.1%
Tiny	44.3%	0.01	53.2%
		0.05	87.4%
		0.1	N/A

<표 4> 적대적 예제에 대한 인식률

Whisper	Original Speech	Adversarial Examples	
	CER	Attack Level(epsilon)	CER
Medium	23.7%	0.01	61.1%
		0.1	26.5%
		0.3	27.1%
Small	25.2%	0.01	28.2%
		0.1	26.8%
		0.3	40.1%
Tiny	44.3%	0.01	97.3%
		0.1	53.4%
		0.3	47.2%

<표 5> 적대적 공격 성공 예시

Whisper - Small	Result
Original Speech	빨리 해야 돼요.
Adversarial Example	The party is over
Ground Truth	이거 빨리 해야 돼요

5. 결 론

본 연구에서는 다국적 음성인식 모델인 Whisper의 취약성에 대하여 실험하여 군사적 환경에서의 활용 가능성을 확인하고자 하였다. 이를 위하여 전장소음에 대한 취약성, 잡음에 대한 취약성, 적대적 공격에 대한 취약성으로 구분하여 실험결과를 도출하였으며, 향후 군사적 활용을 위해 개선해야 할 사항은 아래와 같다.

실험 4.1의 전장소음에 대한 취약성 경우 Medium, Small, Tiny 모델 모두 많은 성능의 저하가 일어난 것을 확인하였고 평균 72.4%의 CER을 나타낸 것으로 보아, 성능개선 없이는 전장소음 환경에서는 활용에 어려움이 있음을 시사한다. 하지만, 소음 환경과 같은 특수한 환경에서의 음성 인식률은 모델의 파인튜닝을 통해서 개선할 수 있다. 따라서, 실제 전장소음 데이터를 수집하여 음성인식 모델의 파인튜닝에 사용하여 전장소음에 대한 취약성을 감소시킬 수 있다.

실험 4.2의 잡음에 대한 취약성의 경우, 강도가 약한 잡음에 대해서는 Whisper가 강건한 것으로 확인할 수 있었다. 하지만 인위적으로 강도가 강한 잡음을 추가하였을 때는 성능이 떨어지는 것을 확인할 수 있었다. 하지만, 잡음의 경우에는 주파수 필터링 및 퓨리에 변환 기반의 모델을 시작으로 Wavelet transform[12]과 같은 다양한 모델들이 잡음 제거에 높은 성능을 나타내기 때문에, 잡음이 섞인 데이터의 경우 디노이징 모델을 음성인식 모델과 함께 활용함으로써 잡음에 대한 취약성을 감소시킬 수 있다.

실험 4.3의 적대적 공격에 대한 취약성에서는 모델마다 특정 입실론으로 생성된 적대적 예제에서 취약한 것을 확인할 수 있었다. 적대적 공격은 적대적 학습을 통하여 모델의 취약성을 줄일 수 있다. 하지만, Whisper의 경우 모델에 따라 취약한 입실론 계수가 달라지기 때문에 적대적 학습시 다양한 입실론 계수를 통해 생성된 적대적 샘플을 학습하는 것이 중요하다.

본 연구의 2장에서는 음성인식 모델이 취약성을 보이는 경우에 대하여 관련연구를 조사하였고, 3장에서는 취약성 검증을 위한 실험데이터를 구성였다. 이를 바탕으로 진행한 4장의 실험결과에 따라, 다국적 음성인식 모델인 Whisper를 당장 군에서 도입하기에는 많은 취약점이 존재하는 것을 확인하였다. 하지만 음성인식 모델은 여단급 이상의 AI참모 시스템, 대대급 이하의 실시간 사격통제, 제대별 무선통신(명령하달) 인식률 개선 등 다양한 시스템에 반드시 포함된다. 따라서, 음성인식 모델 도입 시 군사적 환경에서 고려해야 하는 사항과 개선방법을 제시하였으며, 향후 연구 방향은 6장에 기술하였다.

6. 향후 연구

군사적 환경에서 다국적 음성인식 모델인 Whisper의 취약점을 개선하기 위해 Whisper 모델을 파인튜닝 기법을 활용하고자 한다. 이때 소음 및 잡음, 그리고 적대적 공격에 대한 강건한 모델로 개선하기 위해서 전장 소음, 잡음, 적대적 예제를 포함한 데이터를 동시에 학습시켜 파인튜닝과 적대적 학습을 진행하고자 한다. 또한, 소음 및 잡음이 포함된 음성데이터에 대해서 더욱 강건하게 만들기 위하여 디노이징모델을 함께 적용해 보고자 하며, 이를 통해 개선된 모델의 군사적 환경에서 사용 가능성을 재평가하고자 한다.

참고문헌

[1] Malik, Mishaim, et al. "Automatic speech recognition: a survey." Multimedia Tools and Applications 80,

9411-9457, 2021.

[2] 김영진, 차현종, 강아름. "음성 데이터 품질이 음성 인식 모델에 미치는 영향 연구." 한국컴퓨터정보학회논문지 29.1, 41-49, 2024.

[3] Kwon, Hyun, et al. "Selective audio adversarial example in evasion attack on speech recognition system." IEEE Transactions on Information Forensics and Security 15 (2019): 526-538.

[4] Ko, Kyoungmin, et al. "Multi-targeted audio adversarial example for use against speech recognition systems." Computers & Security 128, 103168, 2023.

[5] Hannun, Awni, et al. "Deep speech: Scaling up end-to-end speech recognition." arXiv preprint arXiv:1412.5567, 2014.

[6] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30, 2017.

[7] Radford, Alec, et al. "Robust speech recognition via large-scale weak supervision." International Conference on Machine Learning. PMLR, 2023.

[8] Baevski, Alexei, et al. "wav2vec 2.0: A framework for self-supervised learning of speech representations." Advances in neural information processing systems 33, 12449-12460, 2020.

[9] C. Oh, et al. "Building robust Korean speech recognition model by fine-tuning large pretrained model." Phonetics and Speech Sciences 15.3, 75-82 2023.

[10] Dua et al. "Noise robust automatic speech recognition: review and analysis." International Journal of Speech Technology 26.2, 475-519, 2023.

[11] Zhang, Chaoning, et al. "A survey on universal adversarial attack." arXiv preprint arXiv:2103.01498 2021.

[12] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," biometrika, vol. 81, no. 3, pp. 425-455, 1994.

〔저자소개〕

원 엘 림 (El-im Won)
2023년 5월~현재 육군 지능정보기술단
2015년 2월 성신여자대학교 학사
email : el1101@naver.com

나 성 중 (Seong-jung Na)
2015년 2월 육군사관학교 학사
2023년 3월 美 해군대학원 공학석사
2023년 3월 국방대학교 공학석사
2024년 1월 ~ 현재
국방대학교 군사운영분석 박사과정
email : sjna0822@gmail.com

고 영 진 (Young-jin Ko)
2015년 2월 육군사관학교 학사
2021년 6월 미 콜로라도 볼더대 이학석사
2023년 8월~ 육군사관학교
컴퓨터과학과 교수
email : kyj020792@gmail.com