



# I 팀 소개



Where AI & Defense Connect



**임기성**

공군 정보체계관리단 AI/데이터개발팀

한양대학교 ERICA 컴퓨터학부

NIA AI 데이터 구축 정부과제 프로젝트 경험



**권민재**

공군 정보체계관리단 정보화지원팀

Princeton Computer Science 전공

국제 수학 올림피아드(IMO)  
국가대표 은메달(2020년)



**김도연**

공군 항공안전단 정보체계반

서울대학교 컴퓨터공학부

공군 AI 해커톤 본선 진출 경험



**이유한**

공군 항공안전단 정보체계반

서울대학교 전기정보공학부

공군 AI 해커톤 본선 진출 경험

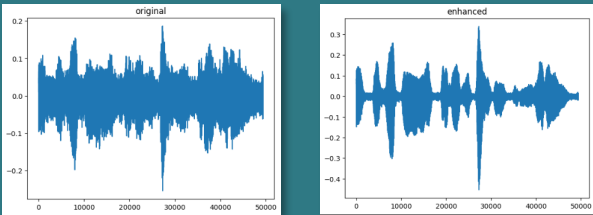
# II 발표 내용\_ [평가 항목] 문제 이해도

## 문제 개요

음성 오디오 데이터를 텍스트로 전사하는 문제

### 전처리

#### 노이즈 제거

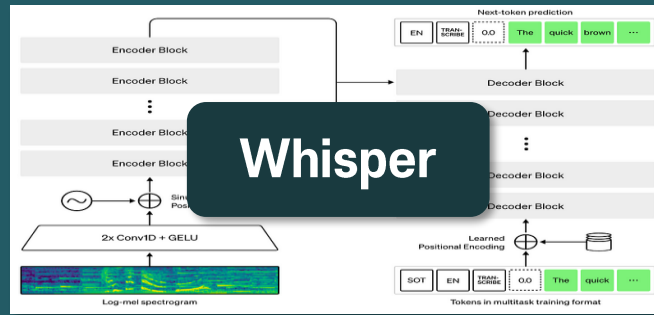


#### EDA를 통한 전처리

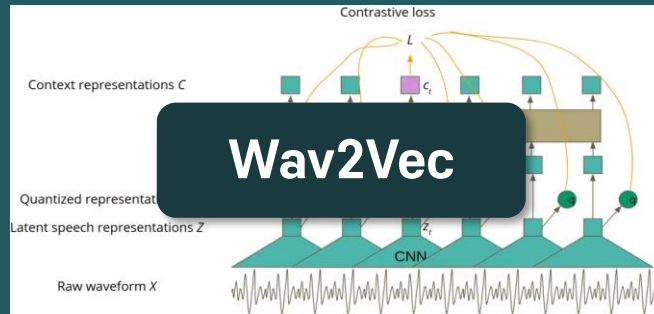
화자의 성별에 따른 분리

Sampling Rate 이상치 제거

### ASR 모델



Whisper



Wav2Vec

### 후처리

#### 중복 Output 처리

같은 단어가 반복되는 Output 존재

기초생활수급자로 등록하게 속도 가르쳐 줘.  
네. 네. 네. 네. 네. 네. 네. 네. 네. 네.  
중식중 영상 맨 뒤 화면에 올려 뒤.

#### 잘못 인식된 문장 처리

음성을 잘못 인식하여  
맥락이 맞지 않는 문장 존재

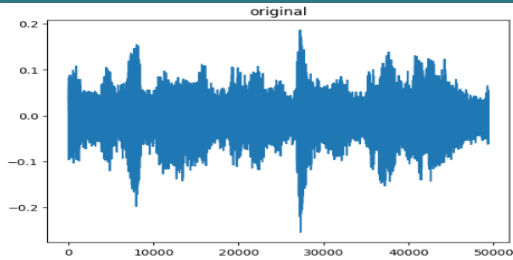
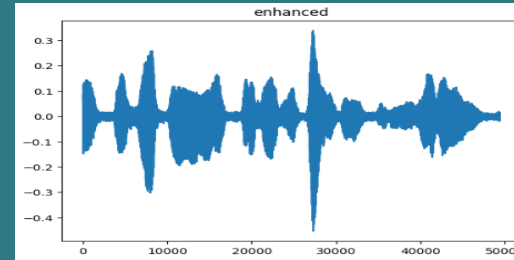
케이씨 통신사 위치 검색해 드릴게요.  
금요일 푸드코트 기온 알려 줬으면 해.  
지금 인기 찾기 못해.

# II 발표 내용\_ [평가 항목] 데이터 활용



### 노이즈 제거

새소리, 전파 노이즈, 세탁기 소리, 노래 소리 등 다양한 노이즈 존재

SOTA 모델은 영어 발성에 맞춰 Denoising 되어 성능 저하 판단  
한국어 음성으로 Pre-train 된 Denoising 모델 활용. Score 상승 (+0.03)

### 중복 데이터 후처리

기초생활수급자로 등록하게 속도 가르쳐 줘  
네. 네. 네. 네. 네. 네. 네. 네. 네. 네.  
중식중 영상 맨 뒤 화면에 흘러 댐.

출력 데이터에 동일한 문구를 여러 번 반복하는 현상 확인

3번 이상 반복되는 substring을 제거하는 PostProcessing 작업  
Score 상승 미비 (+0.001)

### 인식 오류 데이터 후처리

케이씨 통신사 위치 검색해 드릴게요.  
금요일 푸드코트 기온 알려 줬으면 해.  
지금 인기 찾기 못해.

sym\_spell 라이브러리 사용.  
단어 기준 Correction 시도.

한국어 맞춤법 교정 모델 사용.  
오히려 Score 하락 (-0.011)

### 화자의 성별에 따른 분리

여성 데이터 수: 11284 CER: 0.24  
남성 데이터 수: 2716 CER: 0.17

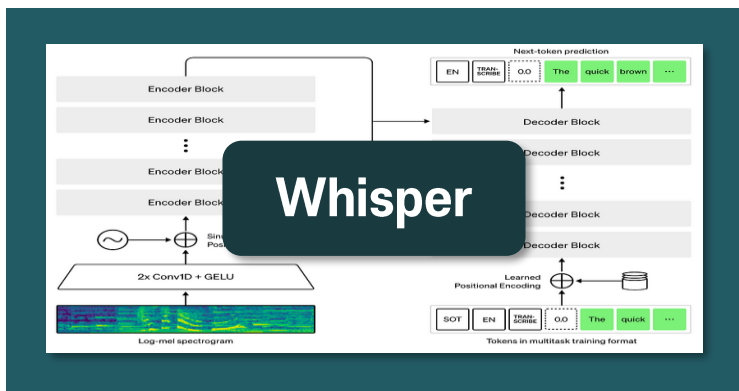
화자 성별 분석  
Classification 모델 활용

성별에 따른 데이터 분포 비교  
성별간 데이터 불균형 확인

성별에 따른 CER 분석  
남성이 여성에 비해 CER이 낮음

여성 데이터를 남성 음역대로 변환,  
성별간 모델 분리 학습 등 시도

# II 발표 내용\_ [평가 항목] 모델 완성도



대표적인 ASR모델 2개 Wav2Vec, Whisper 직접 Inference 하여 성능 비교

Wav2Vec CER  
0.85

Whisper CER  
0.80

Wav2Vec 보다 Whisper 성능이 우월하다고 판단

적은 데이터와 한정된 자원의 상황  
한국어 pretrained 모델 활용

[huggingface]  
seastar105/whisper-small-ko-zeroth

Model	Whisper small
Batch Size	8
Steps	6000
Learning Rate	1e-5
Optimizer	AdamW

## 모델 Input 구조에 맞춰 데이터 변경

Whisper 모델의 입력 처리 기준은 30초.  
그러나 대부분의 데이터가 짧은 데이터 (약 4초) 이므로  
여러 데이터를 30초에 맞춰 병합하여 학습 진행

한번에 학습하는 데이터가 많아져 학습 효율 증가를 기대

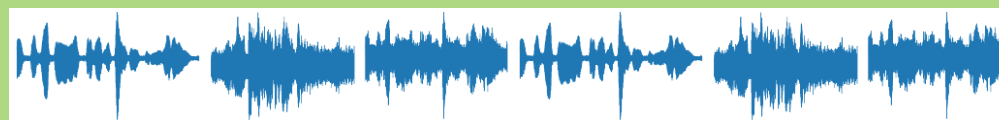
실제 Validation 성능 감소. 음성 길이 분포 차이로 인한  
Overfitting 으로 추정

## Input



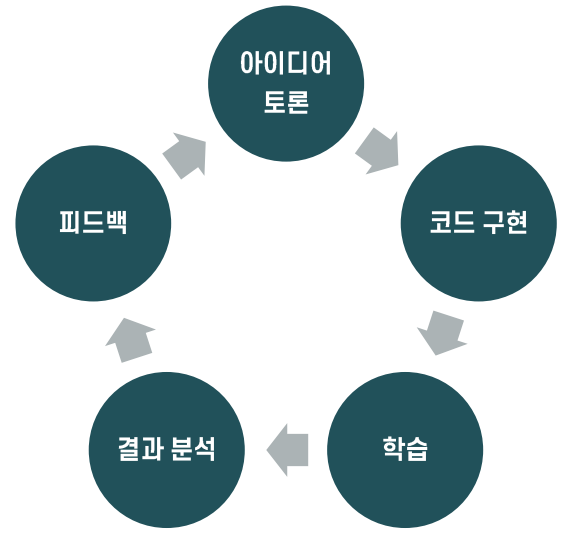
30s

## Input



30s

# II 발표 내용\_ [평가 항목] 수행 프로세스



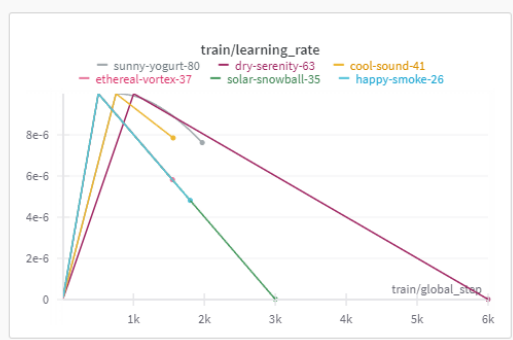
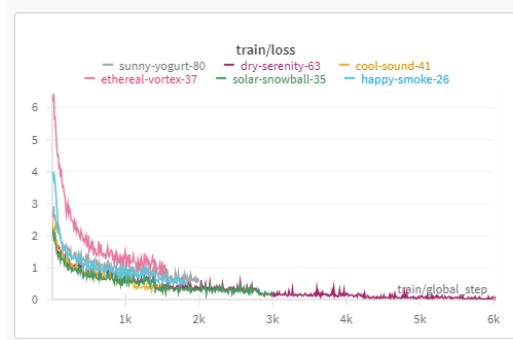
## 역할 분배

각 프로세스에 적합한 인원이 해당 역할 수행

임기성	프로젝트 총괄, DevOps, 발표
권민재	도메인 지식 조사
김도연	전처리 프로세스 구축
이유한	실험용 코드 구축

## 학습 관리

WandB 를 사용하여 실시간으로 모델 학습 현황 확인 및 검증



### 1일차

베이스라인 구축    모델 비교

### 2일차

Whisper Base 모델 학습    Pretrained Model 학습

### 3일차

제출 코드 구축

노이즈 제거

성별 간 데이터 분석

모델 Input 구조 맞춤

Output 데이터 후처리

발표 자료 제작



# 2023 국방 AI 경진대회



## Where AI & Defense Connect

### 소감 한마디

이번 2023 MAICON은 문제 해결을 위한 다양한 아이디어를 함께 고민하는 과정이었습니다.  
많은 방법들을 구상하고 시도해 보았으나, 결과적으로 시도한 방법들이 실제 문제를 개선하지는 못해서 아쉽습니다.

익숙하지 않은 Audio 분야 문제를 다양한 관점에서 바라보고, 여러 아이디어를 시도하며 많은 것들을 배웠습니다.  
결과적으로 좋은 성능을 내지는 못했지만, 다양한 시도를 해보고 실패했기에 좋은 경험이 되었다고 생각합니다.