

2023

군장병 부문

국방 AI 경진대회

결과물 발표평가



팀명 | G-PT



대한민국 국방부
Ministry of National Defense



과학기술정보통신부
Ministry of Science and ICT



대위 김동휘



공군 정보체계관리단
체계개발대 AI/데이터개발팀장
Denoising 및 데이터 후처리 담당



대위 김동민

국직 사이버작전사령부
3작전단 무기체계안전대
안전관리담당
EDA 및 모델 학습 담당



대위(진) 전상범

공군 정보체계관리단
체계개발대 AI/데이터개발팀
시체계기술담당
Baseline 선정 및 모델 학습 담당



중위 이효진

공군 정보체계관리단
체계개발대 AI/데이터개발팀
빅데이터분석담당
EDA 및 Augmentation 담당

과제 내용

음성인식 기술을 활용한 효과적인 지휘관의 전술 명령 전달

ASR(Automatic Speech Recognition)



- 음성(Speech)을 글자(Text)로 바꾸는 기술
- 개발환경과 task에 맞는 Baseline 모델 선정 필요
- 주요 고려사항
 - 사람이 발화하는 주파수대역(100~250Hz)을 잘 인식해야 한다.
 - Noise 제거가 중요.
- 모델 결과의 데이터 후처리를 통해, 이해 가능한 텍스트 만드는 것이 중요.

문제 해결 전략

Baseline 모델 선정

- ASR 과제에서 SOTA 모델 : Whisper, Wav2Vec2.0 조사
- 한국어로 pre-training된 모델을 찾아볼 것
- 제한된 자원과 시간 내에 문제 해결할 수 있는 모델 크기와 사용법 고려

GPU : A100 20GB

데이터 활용

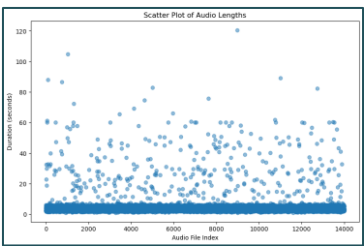
- EDA(Exploratory Data Analysis)를 통한 데이터 이해
- Speech Denoising의 SOTA 모델 사용하여 전처리
- 화자의 발음을 분명하게 하는 Data Augmentation 시도

모델 완성도 향상

- Wandb 라이브러리를 활용하여 모델 학습 관리
- 비슷한 톤의 데이터 예측을 위한 epoch 증가.
- 모델의 결과에 데이터 후처리 진행

EDA(탐색적 데이터 분석)

1 Audio 음성길이 분포 확인

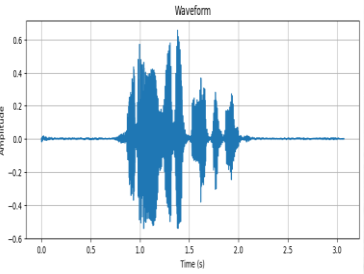


- 특정 audio 파일의 길이가 긴 것을 확인
- Whisper 모델 특성 상, 30초 이상 Audio는 학습/예측에 제한



- 30초 이상 Audio는 학습에서 제외
- 추론 시, 긴 audio는 분할해서 추론하는 pipeline 이용

2 Audio 경향 분석

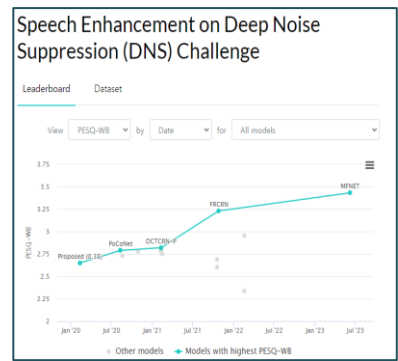


- 비슷한 톤의 발화 음성이 다수(대부분 여성 발화)
- Train data에서 unique한 text는 4447개. 같은 text에 대해 다른 noise로 Audio 구성

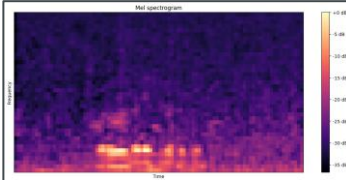


- Train data의 집중된 학습 가능
- 여성 음성 스펙트럼에 강조를 주는 Augmentation 시도

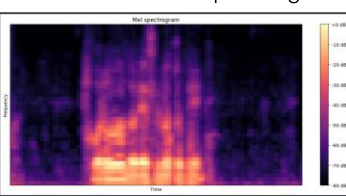
Denoising 전처리



<Denoise 전 mel-spectrogram>



<Denoise 후 mel-spectrogram>

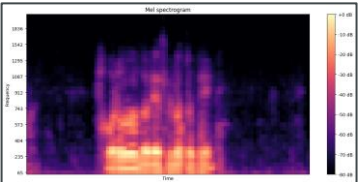


- paperswithcode에서 Speech Denoising SOTA 모델 적용
 - FRCRN 음성 잡음 감소 모델(SOTA) 활용
 - 다양한 소음 환경에서의 단일 채널 음성 지능형 소음 감소 모델 알고리즘
 - Deep Complex CRN 구조 사용
- Library가 잘 되어있어, 빠른 시간내에 적용 가능
- Denoising에서 좋은 성능을 확인 가능

Data Augmentation

1 사람 주파수 강조

<Aug 후 mel-spectrogram>



- 발음이 불분명한 오디오가 다수
- 음성 주파수 대역을 강조하는 전처리 진행
- scipy.signal.lfilter 사용

2 VAD(Voice Activity Detection) 시도

- VAD 처리 후, Denoising하는 선행연구 존재
- 하지만 본 과제에서 음성이 지워지는 현상 발생



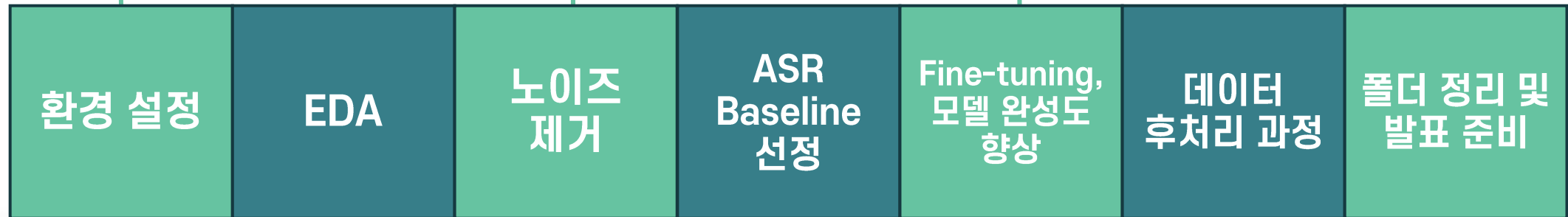
- Augmentation 처리할 경우, 데이터 추론에서 성능 저하.
- Denoising 전처리만 진행하여 학습 및 추론

* Mel-Spectrogram: 사람의 음성 주파수 정보를 표현한 그래프

- 각자 역할분담에 따른 환경설정
- Anaconda, vscode, Libray 등 필요한 환경 설치 및 설정

- FRCRN Denoising 모델 (SOTA) 활용, 소음 제거

- Denoising 처리 및 pipeline 적용
→ Noise 및 긴 길이 데이터 처리 가능
- 다양한 epoch으로 training 및 결과 관리



- 음성 길이와 경향성 파악
→ 음성의 톤, 발화자 성별, text의 내용 등

- “whisper-medium-ko-zeroth” 선정
→ 한국어 pre-training된 모델
→ 제한된 환경에서 학습 가능

- 특수기호 및 반복 어절 삭제



2023 국방 AI 경진대회



Where AI & Defense Connect

소감 한마디

- 제한된 환경속에서 최대의 성능을 위한 Baseline 모델 선택과 데이터 분석의 중요성을 알 수 있었다.
- ASR 분야에서 변환된 텍스트를 자연어 모델로 정확도를 향상시킬 수 있을 것이라고 생각된다.
- 좋은 대회 주최/운영해주신 모든 관계자분들께 감사드립니다. ♡