



# I 팀 소개



**대위(진) 윤인수**

제17전투비행단 항공작전전대  
통합체계운영대대

<팀장>



**대위(진) 정운영**

작전정보통신단  
체계개발실

<EDA 및 데이터 전처리>



**병장 신지훈**

제10전투비행단 수송대대  
운영통제실 배차실

<데이터 전처리 및  
자료 정리>

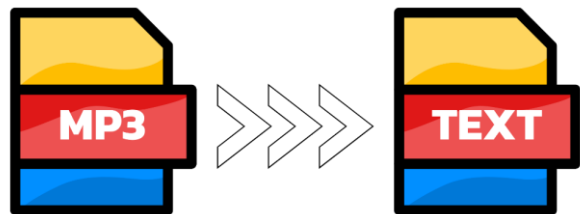


**일병 서덕훈**

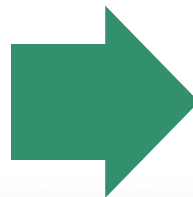
제16전투비행단 정보통신대대  
사이버중대 정보체계반

<모델 개발 및 자원 관리>

## II 발표 내용\_ [평가 항목] 문제 이해도

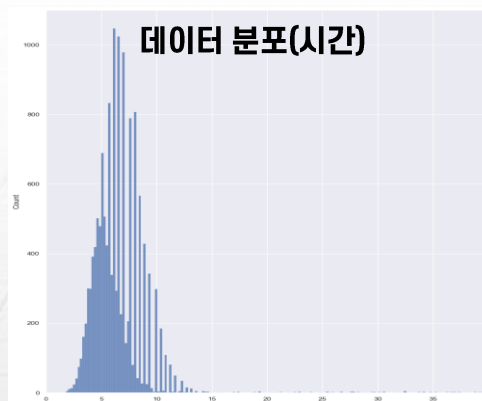


“주어진 Audio(Speech) data를  
Text data로 변환하는 Task”



### [특이사항]

- 14,000개의 한국어 Audio Data.
- Audio 파일의 길이가 다양함.
- 임의로 추가된 Noise가 큼.  
(소음환경)



### Augmentation

- Gain
- Time Stretch
- Deal with Noise

### Speech Enhancement

- Frequency Filter
- Spectral Gating
- Learning Model

### Speech-To-Text Model

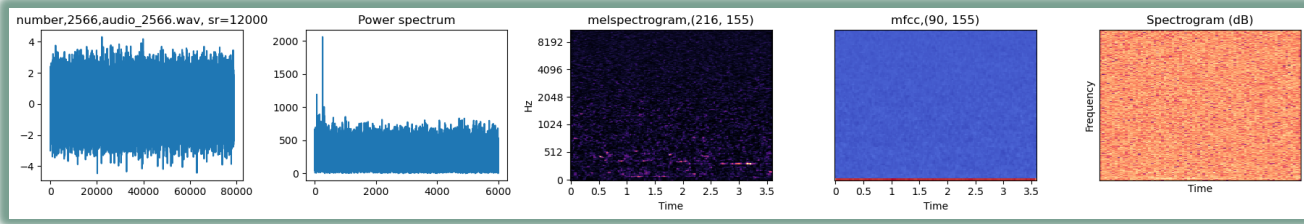
- Openai / Whisper
- Facebook / Wav2Vec



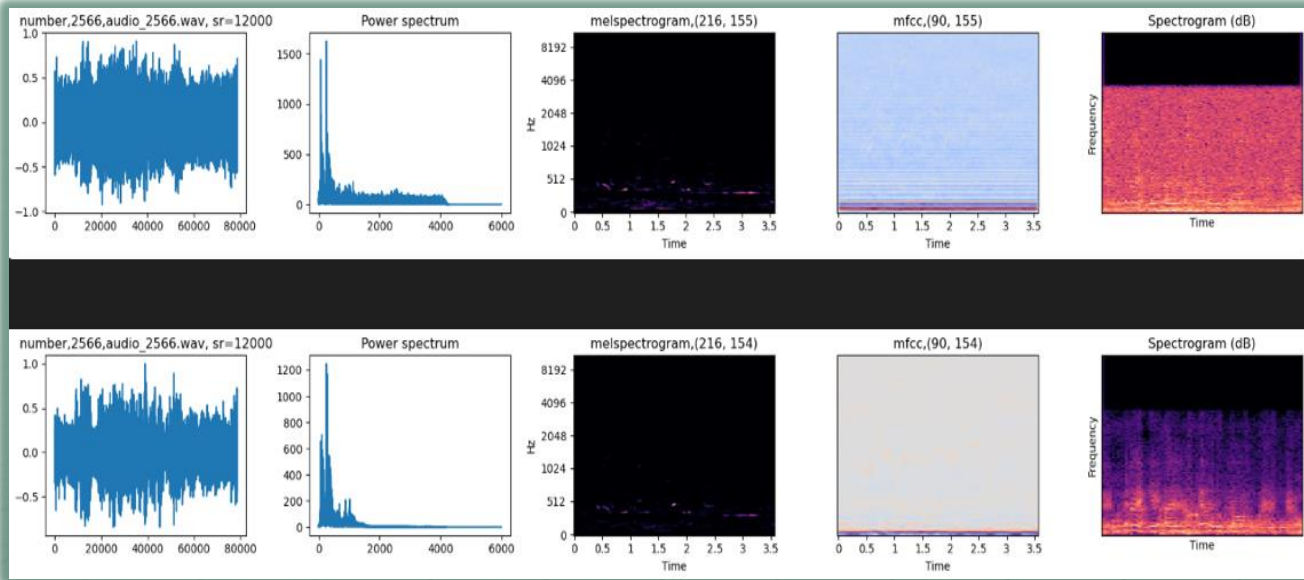
# II 발표 내용\_ [평가 항목] 데이터 활용



## Case 1) Audio data에 Augmentation 및 노이즈 추가



## Case 2) Audio data에 Denoise model 적용



## Case 3) Train.csv data를 Text-To-Speech(TTS)의 자료로 사용

	filenames	text
0	audio_0.wav	이미 달리 환을 어제보다 올랐어?
1	audio_1.wav	뉴스 보게 텔레비전 좀 틀어 놔라.
2	audio_2.wav	편의시설 뭐가 있나 사이트 검색해서 알려드리겠습니다.
3	audio_3.wav	빨리 원무과까지 길 안내 좀 해 주게.
4	audio_4.wav	목적지 선릉까지 지하철 몇 번 갈아타야 하는지 알려 줘.
...	...	...
13995	audio_13995.wav	시월 삼십일 마지막 수술 몇 시야?
13996	audio_13996.wav	병원 진찰 언제로 예약했었지?
13997	audio_13997.wav	음악 시간 준비물로 뭐가 있는지 알려줘요.
13998	audio_13998.wav	일호선을 얼마나 타고 가야 하는지 안내해 줬으면 해.
13999	audio_13999.wav	여기서 바로 카드로 결제해 줄 수 있어?

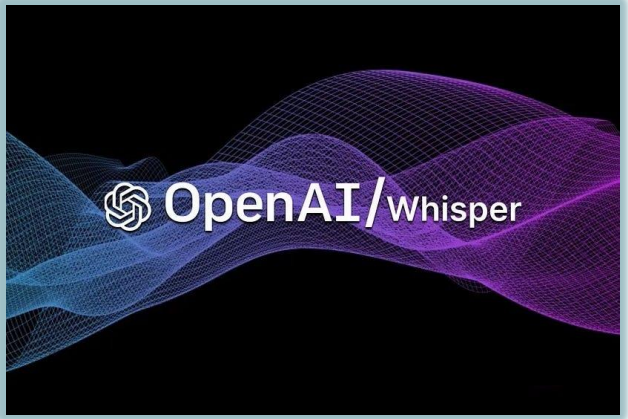
“뉴스 보게 텔레비전 좀 틀어 놔라.”



# II 발표 내용\_ [평가 항목] 모델 완성도

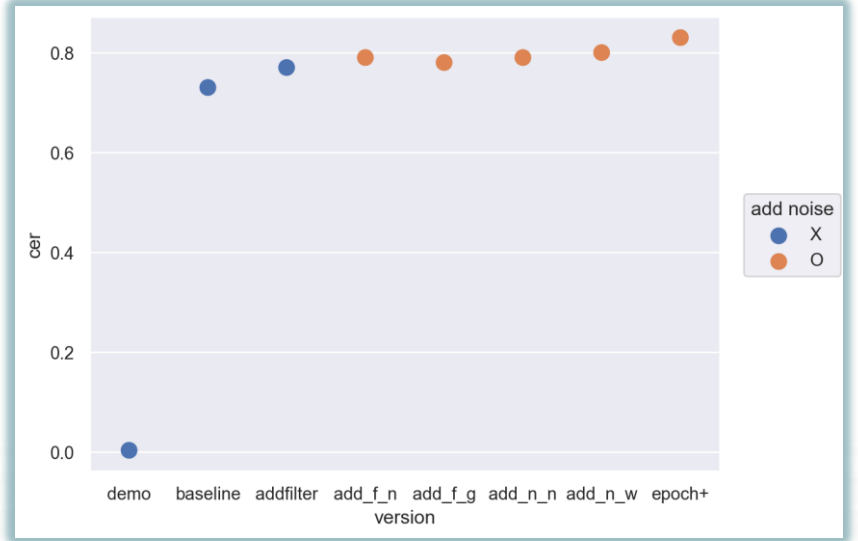


## Whisper-small



Dataset	wav2vec 2.0 Large (no LM)	Whisper Large V2	RER (%)
LibriSpeech Clean	2.7	2.7	0.0
Artie	24.5	6.2	74.7
Common Voice	29.9	9.0	69.9
Fleurs En	14.6	4.4	69.9
Tedlium	10.5	4.0	61.9
CHiME6	65.8	25.5	61.2
VoxPopuli En	17.9	7.3	59.2
CORAAL	35.6	16.2	54.5
AMI IHM	37.0	16.9	54.3
Switchboard	28.3	13.8	51.2
CallHome	34.8	17.6	49.4
WSJ	7.7	3.9	49.4
AMI SDM1	67.6	36.4	46.2
LibriSpeech Other	6.2	5.2	16.1
Average	29.3	12.8	55.2

Table 2. Detailed comparison of effective robustness across various datasets. Although both models perform within 0.1% of each other on LibriSpeech, a zero-shot Whisper model performs much better on other datasets than expected for its LibriSpeech performance and makes 55.2% less errors on average. Results reported in word error rate (WER) for both models after applying our text normalizer.

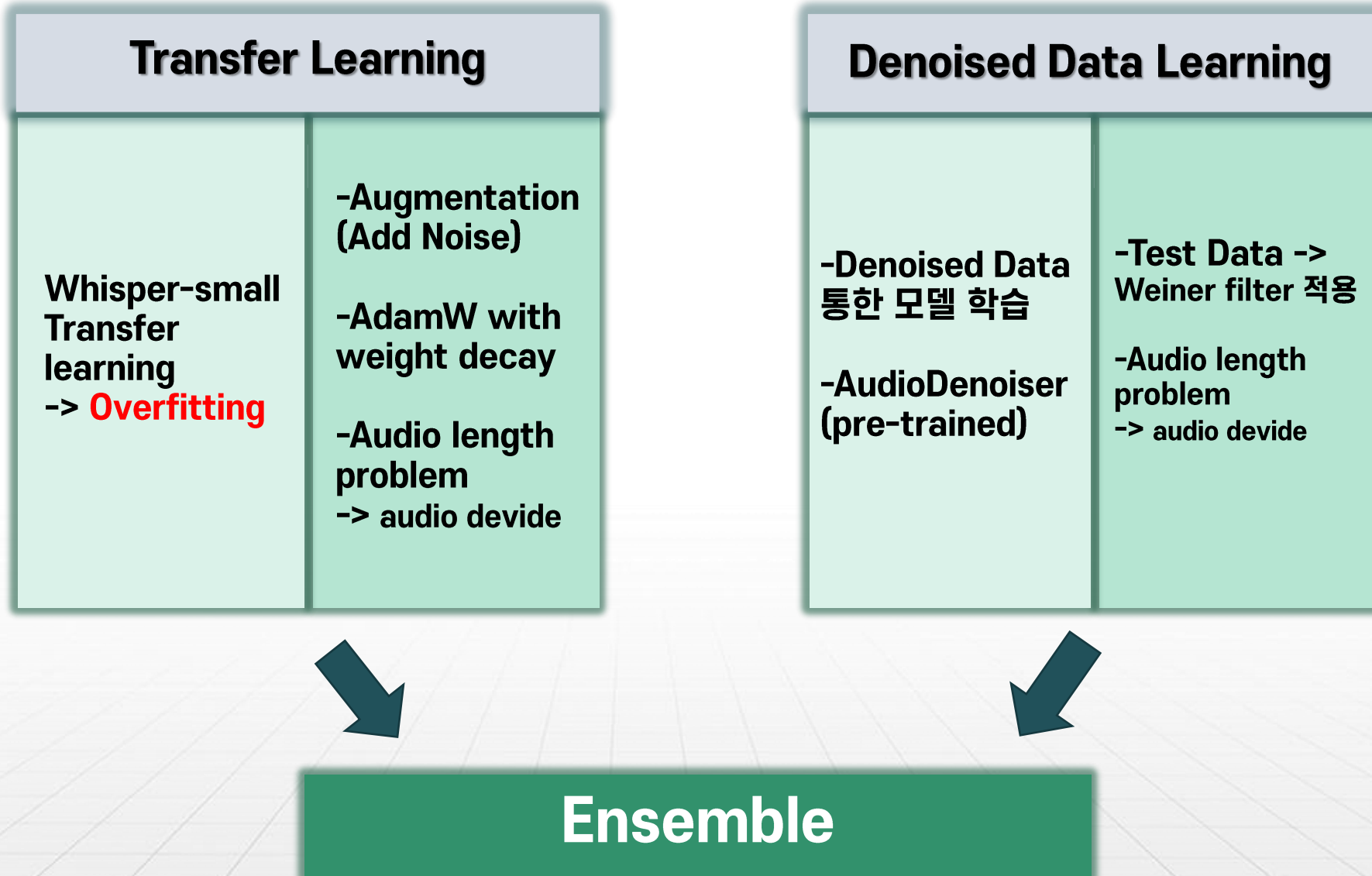


VS

wav2vec2



Robust Speech Recognition via Large-Scale Weak Supervision : [Alec Radford\(2022\)](#), arxiv





## 2023 국방 AI 경진대회



# Where AI & Defense Connect

소감 한마디

**TTS를 활용을 해보았으면 하는 아쉬움이 남는다. (다양한 환경에서만 동작?!)  
지휘관 음성 명령 전달에서 발전하여,  
전투기 조종사 <-> 관제탑 간의 통신 보조, 화자 분리를 적용한  
회의록 작성 보조 등에 적용해 볼 수 있을 것 같다.**