

# Woojin Choi

## Topic : English & Korean Automatic Speech Recognition for Understanding Voice-Commands

### [Current Task]

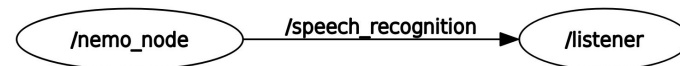
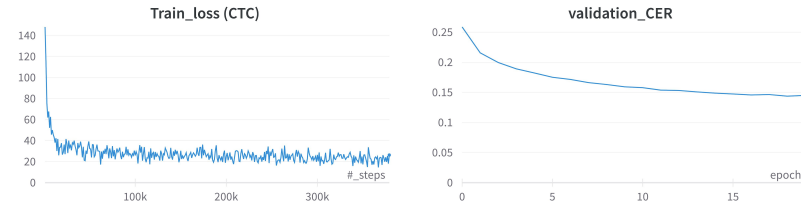
- Build ROS package for offline-Speech Recognition

### [This Week]

- Model Training
  - 1) Korean : Conformer-CTC (~50 Mb), 20 epochs / English : use pre-trained
  - 2) Larger model currently training (~100 Mb, 50 epochs)
- ROS Package “Nemo-ASR”
  - Set Language on node start (“en”, “ko”)
  - Record & transcribe speech for fixed-time frame (ex. 5 seconds)
  - Publish ASR result as topic
- Results
  - Korean ASR : **CER 15.9 %** (tested on KsponSpeech-eval)
  - Integration w/ NLU packages : Replaced Google-speech node
- Discussion
  - Out-of-Vocabulary 한글 characters, digits, punctuations, etc.
  - (ex) “30 미터 앞으로” -> “삼십 미터 앞으로”
  - Test w/ robot : **IRL or simulation (Quadruped)** ?

### [Next Week]

- 1) Demonstration : Legged robot / Simulation
- 2) Replace best-performing model
- 3) Report & Final presentation



```
[INFO] [1675943914.069964]: nemo-recognizer started.
[y/n] 'y' to record for 5 seconds / 'n' for shutdown
y
[INFO] [1675943919.153138]: Recording audio..
Transcribing: 100%|
[INFO] [1675943924.318165]: Transcription : bring me the ball from the table
y
[INFO] [1675944000.005075]: Recording audio..
Transcribing: 100%|
[INFO] [1675944005.159273]: Transcription : 책상에서 공을 가져와.
[y/n] 'y' to record for 5 seconds / 'n' for shutdown
```

```
cwwojin@ubuntu:~/catkin_ws$ roscpp_nemo_asr_listener.py name:=listener
[INFO] [1675943924.318841]: /listener heard : bring me the ball from the table
[INFO] [1675943935.965743]: /listener heard : move to the front
[INFO] [1675943946.089187]: /listener heard : go to the rightmost chair
[INFO] [1675944005.159617]: /listener heard : 책상에서 공을 가져와.
[INFO] [1675944016.917407]: /listener heard : 앞으로 이동해.
[INFO] [1675944031.518762]: /listener heard : 가장 오른쪽 의자로 이동해.
```

Task / Week		(1/2 ~ 8)	(1/9 ~ 15)	(1/16 ~ 20)	(1/25 ~ 29)	(1/30 ~ 2/5)	(2/6 ~ 12)	(2/13 ~ 19)
ROS-Vosk	TRUE							
ROS / NeMo tutorial	TRUE							
Dataset gathering	TRUE							
ASR model training	TRUE							
Build ROS package	TRUE							
Integrate w/ SGGNet	FALSE							
Demo - robot / simulation	FALSE							
Report & presentation	FALSE							

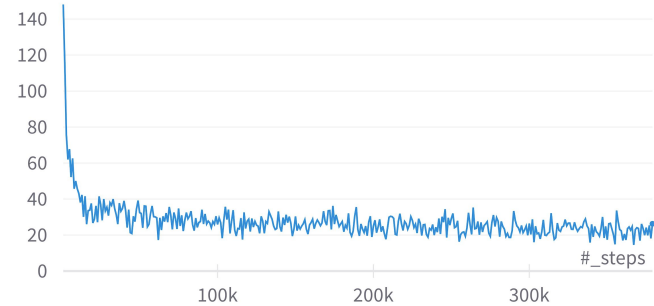
figures

- 1. results : loss / CER graph
- 2. Inference examples
- 3. ROS package interface (run, key-input, results printed)
- 4. ROS graph (demo)

```
[INFO] [1675943914.069964]: nemo-recognizer started.  
[y/n] 'y' to record for 5 seconds / 'n' for shutdown  
y  
[INFO] [1675943919.153138]: Recording audio..  
Transcribing: 100%|  
[INFO] [1675943924.318165]: Transcription : bring me the ball from the table  
y  
[INFO] [1675944000.005075]: Recording audio..  
Transcribing: 100%|  
[INFO] [1675944005.159273]: Transcription : 책상에서 공을 가져와.  
[y/n] 'y' to record for 5 seconds / 'n' for shutdown  
y  
cwwojin@ubuntu:~/catkin_ws$ rosrn nemo_asr listener.py __name:=listener  
[INFO] [1675943924.318841]: /listener heard : bring me the ball from the table  
[INFO] [1675943935.965743]: /listener heard : move to the front  
[INFO] [1675943946.089187]: /listener heard : go to the rightmost chair  
[INFO] [1675944005.159617]: /listener heard : 책상에서 공을 가져와.  
[INFO] [1675944016.917407]: /listener heard : 앞으로 이동해.  
[INFO] [1675944031.518762]: /listener heard : 가장 오른쪽 의자로 이동해.
```



Train\_loss (CTC)



validation\_CER

