# CS310 Assignment 3

Tom Ko

# Objectives

- Go through the whole process of training a Chinese ASR system.
- Apply data augmentation.

# Our test set

- In the lab, you use the aishell test set.
- In this assignment, everyone of you have to contribute to the test set.

# Recording your own voice

- Everyone of you has to provide 10 recordings of your own voice.
- Each recording has to be at least 7 sec and contains at least 30 Chinese characters.
  - E.g. 最近情况好了很多 真的太好了 你知不知道学校开始要我们回校了 我很期待
- You have to label your own data.
- Think about the most common sentences that you speak in your daily life.

# File names

- We have 27 students in the class, thus we will have at least 270 recordings as the test set. (The TAs will also contribute to the test set.)
- All recordings have to be *.wav files.
- Please record with 16000 Hz sampling rate.
- Name your files "SUST{studentID}_0{uttid}.wav" where studentID is your student id and uttid is the utterance id ranging from 01-10. E.g. SUST11612341_003.wav is the third utterance of student 11612341.

# Labeling format

- Please label your data as the Kaldi format.
- Look at the aishell data to checkout the format.
- You have to prepare a file named “text”. The content is like
  - SUST11612341_001 今天天氣怎
  - SUST11612341_002 我先回家了
  - SUST11612341_003 肚子餓了 我去吃飯
  - . . . . .

# wav.scp

- You have to prepare a file named “wav.scp”. The content is like
  - SUST11612341_001 /data/cs310/XXX/assignment3/wav/SUST11612341_001.wav
  - SUST11612341_002 /data/cs310/XXX/assignment3/wav/SUST11612341_002.wav
  - SUST11612341_003 /data/cs310/XXX/assignment3/wav/SUST11612341_003.wav
  - . . . . . .

# utt2spk

- You have to prepare a file named “utt2spk”. The content is like
  - SUST11612341_001 SUST11612341
  - SUST11612341_002 SUST11612341
  - SUST11612341_003 SUST11612341
  - . . . . . .
- After creating your utt2spk file, the spk2utt file can be generated by

  *utils/utt2spk_to_spk2utt.pl* utt2spk > spk2utt

# Test with your own data

- You can use linux command “soxi” to check the audio format of your files.
- After you have prepared your data, you can decode them with the aishell model which you obtained in your lab.
- In order to do that, you have to:
  - Pack your data into kaldi format (wav.scp, utt2spk, spk2utt, text). Check them out in aishell data
  - Extract MFCC of your data
  - Decode them with your model

# Data augmentation

- Your score for this assignment will be proportional to the accuracy on the test set given by your model.
- If the test set is recorded in noisy environment, you are suggested to apply data augmentation techniques.

# Adding noise to your training data

- In Kaldi, you can use the noise adding script in this way:

```
utils/data/get_utt2dur.sh data/train
mv data/train/utt2dur data/train/reco2dur

musan_root=/home/cseadmin/source/musan
 # Prepare the MUSAN corpus, which consists of music, speech, and noise
steps/data/make_musan.sh --sampling-rate 16000 $musan_root data

# Augment with musan_noise
steps/data/augment_data_dir.py --utt-suffix "noise" --fg-interval 1 --fg-snrs "15:10:5:0"
--fg-noise-dir "data/musan_noise" data/train data/train_noise

# Combine noise and the original data
utils/combine_data.sh data/train_combined data/train data/train_noise
```

# Training

- The setting just shown is an example, they are not guaranteed to improve accuracy in you test set. You are suggested to look at the option of the .py and tune the parameters.
- data/train is your original training data and data/train_noise is the augmented data.
- Please look at data/train_noise/wav.scp and see how kaldi operate in noise addition.
- data/train and data/train_noise should be combined to form a new folder, you should use it to train the models.
- After generating the noised folder, remember to do MFCC extraction.
- Please refer to kaldi/egs/voxceleb/v2/run.sh to understand how this script are used.

# Training

- In the lab, you are suggested to stop at training stage “tri2”.
- In this assignment, you can do further training after stage “tri2”.
- You should test the models with your own voice and submit the model which you think is the most accurate.

# Submission

- You have to place all the necessary files in /data/cs310/XXX/assignment3/, where XXX is your login name
- You should have prepared 3 subfolders: “wav”, “data” and “tri”
- In the “wav” folder, it should contain the 10 *.wav files, all of them should be correctly named.
- In the “data” folder, it should contain the 4 Kaldi data files: text, wav.scp, utt2spk and spk2utt.
- For the “tri” folder, it should be your experiment folder, e.g. tri2 or tri5a, please make sure you have run the mkgraph program and your experiment folder should include the “graph” folder”.

# Grading

- Your score for this assignment will be proportional to the accuracy on the test set given by your model.
- If your model obtain an accuracy of 90%, your score for this assignment is 90