Project Phase II: Urdu Speech Recognition

Instructions:

- The aim of this project is to give you a hands-on with the complete pipeline of a real-life speech processing application.
- You will use the KALDI Automatic Speech Recognition Toolkit to develop Urdu Speech Recognition.
- **Carefully read the submission instructions, plagiarism and late days policy at the end of assignment.**
- Deadline to submit phase I is: **Sunday, 18th April 2021.**

Resources:

- *rename.py*: All of you have recorded the same sentences and saved them in a folder using the same naming convention. In order to differentiate your recordings from that of your partners', this script will rename your recordings and append a speaker ID (your roll number) with it. It will also add the new utterance IDs (recordings names) in PRUS.txt. A sample working of this script is shown below:

Input	نیلم ـے سالگرہ پر ہیڈ سیسموگراف اسود قریشی کے ماتھے پر اینٹھن اور غم کی آتشیں رو محسوس کی
Output	نیلم ـے سالگرہ پر ہیڈ سیسموگراف اسود قریشی کے ماتھے پر اینٹھن اور غم کی آتشیں رو 21100000_001 محسوس کی

Similarly,

Input	1.wav
Output	21100000_001.wav

Note: Please make a copy (backup) of your recordings before running this script as it does the in-place renaming.

- *LM1.gz and LM2.gz*: You are provided with two language model files. LM1.gz is trained on the first 600 sentences of the PRUS.txt while LM2.gz is trained on complete (708 sentences) PRUS.txt
- *lexicon.txt*: A dictionary file containing the grapheme-to-phoneme (in CISAMPA) mapping.

Installation:

Follow this excellent tutorial to install KALDI on your systems.

Experimentation:

Single Speaker (Each member of the group is required to do it for his/ her recordings):

Train a GMM based tri-phone acoustic model on the first 600 sentences of your corpus (PRUS.txt) and test it on the rest of 108 sentences of your corpus. Use the LM1.gz while decoding.

Multi-Speaker:

a) Train a GMM based tri-phone acoustic model on the first 600 sentences of each speakers' corpus (combine first 600 sentences of each member to form one large training corpus) and test it on the rest of 108 sentences of each speaker's corpus (combine 108 sentences of each member to form one testing corpus). Use the LM1.gz while decoding.

b) In this part you will do the experimentation in leave-one-out strategy i.e. train a GMM based tri-phone acoustic model on the complete corpus (708 sentences) of n-1 speakers and test it on the complete corpus of remaining 1 speaker e.g. if there are 5 members namely SP1, SP2, SP3, SP4, SP5 in the group, you need to train 5 different models as shown below. Use the LM2.gz while decoding.

Sr. No.	Train				Test
1	SP1	SP2	SP3	SP4	SP5
2	SP5	SP1	SP2	SP3	SP4
3	SP4	SP5	SP1	SP2	SP3
4	SP3	SP4	SP5	SP1	SP2
5	SP2	SP3	SP4	SP5	SP1

Note: If you are doing project alone, please contact Haris Bin Zia by 31st March 2021 to know how you can perform multi-speaker experiments.

Important Links:

- Follow this excellent walkthrough of Kaldi for training GMM models
- We are not exactly revealing the details how to do testing/ decoding in Kaldi, you need to figure this out yourself. We are listing some important links that can give you pointers (feel free to search internet as well)
- https://kaldi-asr.org/doc/data_prep.html
- https://github.com/harisbinzia/GuftaarShanaas

Submission Instructions:

Create a word document, name it as roll_number1_roll_number2 etc. e.g. 21100000_21100001.docx, enter the below-mentioned details in it for every experiment (e.g. a group with 5 members has to report 11 experiments).

Experiment No. 1/2/3 etc.

Experiment Type: Single / Multi Speaker

Train/Test Details:

Best WER: xyz

Create a folder Experiments with multiple subfolder 1/2/3 etc. for each experiment. Place the below-mentioned three files of each experiment in the respective subfolder.

- Model file (.mdl)
- best_wer file
- file mentioned in best_wer file e.g. wer_16_0.5
- per_spk file

Now, zip the Experiments folder and .docx, name it roll_number1_roll_number2 etc. e.g. 21100000_21100001.zip and upload it to LMS.

Only one submission per group is required.

Plagiarism Policy:

All work (e.g. code, experiments, writeup etc.) MUST be done independently. Any plagiarism or cheating of work from others or the internet will be immediately referred to the DC. If you are confused about what constitutes plagiarism, it is YOUR responsibility to consult with the instructor or the TA in a timely manner. No "after the fact" negotiations will be possible. The only way to guarantee that you do not lose marks is "DO NOT LOOK AT ANYONE ELSE'S CODE, WORK OR WRITE-UP NOR DISCUSS IT WITH THEM".

Late Days Policy:

The deadline of the assignment is final. However, in order to accommodate all the 11th hour issues there is a late submission policy i.e. you can submit your assignment within 3 days after the deadline with 25% deduction each day.