Weekly Report (June 17th – June 24th)

Xuesong Yang

1. Data clean up: remove the lines with ‘!’ in the transcription.

Dir of transcription: /home/research/interns/xyang/mispronounced\_words/proc/models/phone/REC-OUTPUT/sm\_train.data

Statistics: There are 1971 lines with ‘!’. Thereafter, 48292 samples left.

Table Schedule of experiments

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Imbalanced? | FeatureSet | ObjectiveFunction | | Completed? | | ResultsDir | |
| 0 | Anastassia’s folds | YES |  | |  | | NO | |  | |
| 1 | Baseline  () | YES, with “!” | 11 phone level features | | F1\_score\_micro | | YES | | Output\_baseline | |
| Unweighted\_kappa | |
| 2 | Control1 | YES, without “!” | 11 phone level features | | F1\_score\_micro | | YES | | Output\_control1 | |
| Unweighted\_kappa | |
| 3 | Control2 | YES, without “!” | Only word identity | | F1\_score\_micro | |  | |  | |
| Unweighted\_kappa | |
|  |  |  |  | |  | |  | |  | |
|  |  |  |  | |  | |  | |  | |

Statistics for word identity

1. Correct words: 44293
2. Error words: 3531
3. Unique word list:
4. Annotations for word error:
   1. The word is labeled as an error if any phone within the word is labeled as error. Now we only consider that the error occurrence of the word is counted as 1 although multiple phone errors exist within the word.
   2. We do not want to identify the words that were not read correctly as long as they are pronounced correctly. If the stimulus says ‘cat’ and the speaker said ‘dog’ we count this as ‘correct’ as long as the word was pronounced correctly.

Exceptions:

1. For training samples of /ZH/, there is only 1 sample. Therefore, temporally ignore such phonemes