Assignment 2

Gengcong Yan - 1009903
ELEC-E5510 - Speech Recognition

November 17, 2021

1 Question 1

1.1 A

Based on the models in hmm-6 folds, test error on *w400.dict* is **36.84%**, whereas the test error on *w150.dict* is **15.44%**, outperforming the original test. That's because *w150.dict* contains only those words that appear in the evaluation set, the range of prediction results is narrowed down to the test set, and the error rate is naturally smaller compared to the previous one.

1.2 B

The test error on *w150.dict* without word loop is **10.53%**, outperforming previous models in Subsection A. Our dataset provides single utterance in the model, which means only one word are presented in a single utterance. But because of the word loops in previous models, the recognition results towards singe utterance can possibly become multiple words, resulting in more errors. Now, the improved model gives only one word prediction in the recognition process, improving the success rate.

2 Question 2

```
# Plot
x=list(range(2,7))
test_err=[17.02,14.39,12.46,9.82,10.53]

plt.plot(x,test_err, label = "test_err")
plt.xlabel("Model number")
plt.ylabel("Error rate")
plt.legend()
plt.title("Error rate in different models")
plt.show()
```

2.1 A

The logarithmic likelihood values reported by HERest during HMM training process shows in Fig.1. Among all the Gaussian models after splitting in iteration, we choose log probability of the best one as the value for plotting. The fact that the values are getting larger proves that the model is getting better during training.

2.2 B

The word error rate using all the models through hmm-2 to hmm-6 shows in Fig.2. During the training process, the prediction error rate of the model keeps decreasing after splitting and re-estimation. Both curves, including the curve in the previous question, show that the accuracy

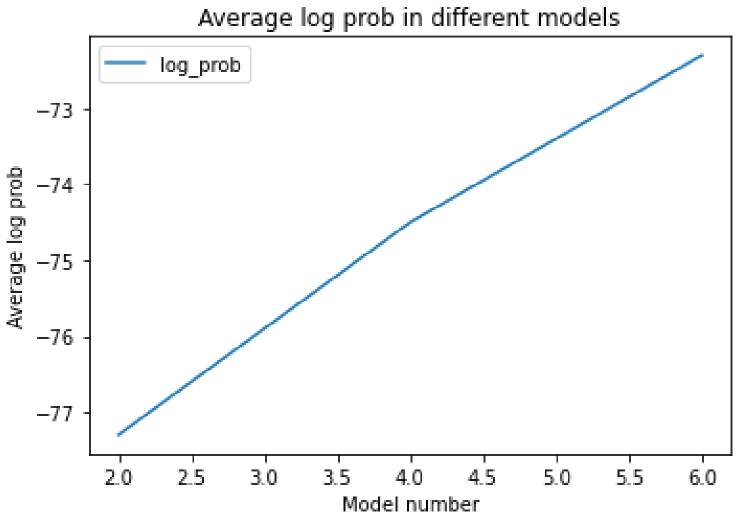


Figure 1: Average log prob in HMM-x models

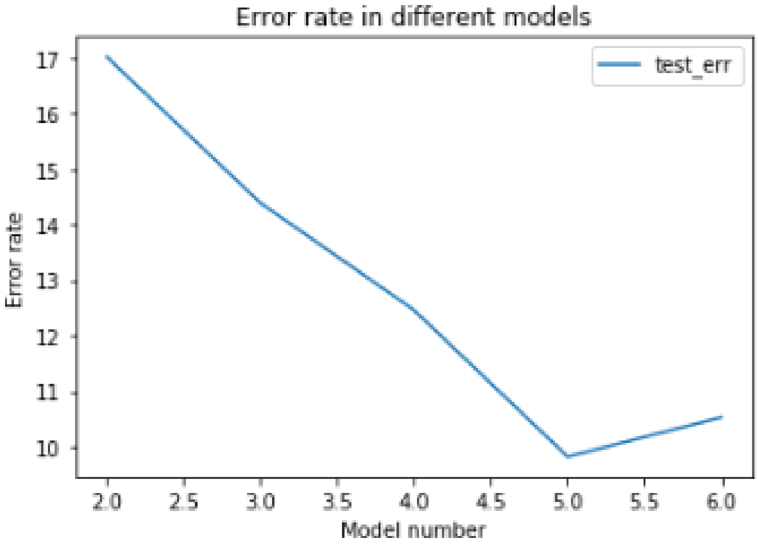
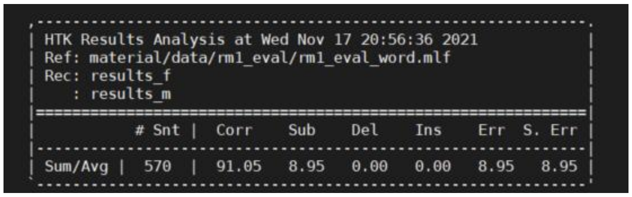


Figure 2: Error rate in HMM-x models

of the model is gradually improving during the training process. Only the information of the log probability curve comes from the internal training, while the information of the error rate curve comes from the evaluation of external data. Finally, as can be seen from the error rate curve in Fig.2, further iterations **do not optimize** the recognition results any more.

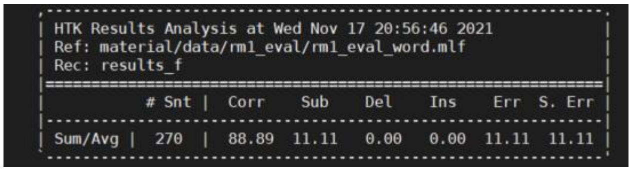
Table 1: Models evaluation

Model	Error rate	Para Num
Female	11.11	313
Male	7.00	766
Combined	8.95	1078



```
,-------------------------------------------------------------.
| HTK Results Analysis at Wed Nov 17 20:56:36 2021            |
| Ref: material/data/rm1_eval/rm1_eval_word.mlf               |
| Rec: results_f                                              |
|    : results_m                                              |
|=============================================================|
|          # Snt |  Corr    Sub    Del    Ins    Err  S. Err |
|-------------------------------------------------------------|
| Sum/Avg |  570  |  91.05   8.95   0.00   0.00   8.95   8.95 |
`-------------------------------------------------------------'
```

Figure 3: Results in combined model



```
,-------------------------------------------------------------.
| HTK Results Analysis at Wed Nov 17 20:56:46 2021            |
| Ref: material/data/rm1_eval/rm1_eval_word.mlf               |
| Rec: results_f                                              |
|=============================================================|
|          # Snt |  Corr    Sub    Del    Ins    Err  S. Err |
|-------------------------------------------------------------|
| Sum/Avg |  270  |  88.89  11.11   0.00   0.00  11.11  11.11 |
`-------------------------------------------------------------'
```

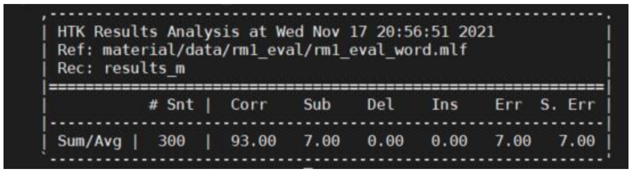
Figure 4: Results in female model

3 Question 3

After building gender dependent acoustic models and training them on data of only the corresponding gender, the following results in Table 1 can be obtained for the recognition error rate of male and female model.

3.1 A

Because the training material is not evenly distributed between the genders, but roughly the proportion of females to males is 2:5. It means for the models based on different genders, they have different amounts of data in training. The more data there is, the more accurate the model's predictions will be. So the results show that the male model has a lower error rate than the female model because it has more data, and the combination of the two has an accuracy between them. The advantage of this scheme is that the models are more targeted to certain samples, we can improve the recognition accuracy if we know in advance the gender of the voice to be recognized and predict it in the corresponding gender model. However, the disadvantages of this are that the requirements for training data are higher, the gender of the sound source in the data needs to be known in advance, and the amount of data in the same case becomes less because the model is trained separately, and we may need more data than before. The detailed information are in following Fig.3, Fig.4 and Fig.5.



```
,-------------------------------------------------------------.
| HTK Results Analysis at Wed Nov 17 20:56:51 2021            |
| Ref: material/data/rm1_eval/rm1_eval_word.mlf               |
| Rec: results_m                                              |
|=============================================================|
|          # Snt |  Corr    Sub    Del    Ins    Err  S. Err |
|-------------------------------------------------------------|
| Sum/Avg |  300  |  93.00   7.00   0.00   0.00   7.00   7.00 |
`-------------------------------------------------------------'
```

Figure 5: Results in male model

3.2 B

If the gender of speakers are male, we perform the recognition on male gender dependent model, whereas if the gender of speakers are female, we can perform the recognition on gender independent model. I think that's the best configuration to use for recognition now.

3.3 C

We use the parameter for the average number of Gaussians in *hmm_train.pl* to adjust for the training. According to the ratio 2:5, the proportion of females to males in data, we set average number of Gaussians in gender dependent models 2.3 and 5.7, respectively. Now we see the parameter numbers in the gender dependent models also reflect this ratio. It's $313 : 766 \approx 2 : 5$.