

# Digital Speech Processing Homework 2 Report

Name：鐘民憲 student ID：B06901017

## 1.Base-Line parameters and Observations

```
### parameters that you can modify
numiters=5 # Number of iterations of training
maxiterinc=4 # Last iter to increase #Gauss on.
numgauss=1 # Initial num-Gauss (must be more than #states=3*phones).
totgauss=5 # Target #Gaussians.
incgauss=$((totgauss-$numgauss)/$maxiterinc) # per-iter increment for #Gauss
realign_iters="1 2 3 4 5";
scale_opts="--transition-scale=1.0 --acoustic-scale=0.1 --self-loop-scale=0.1"
###
```

```
### parameters that you can modify
opt_acwt=0.87
test_beam=15.0
###
```

```
NONSILENCEPHONES
</ForPhones>
<State> 0 <PdfClass> 0 <Transition> 0 0.75 <Transition> 1 0.25 </State>
<State> 1 <PdfClass> 1 <Transition> 1 0.75 <Transition> 2 0.25 </State>
<State> 2 <PdfClass> 2 <Transition> 2 0.75 <Transition> 3 0.25 </State>
<State> 3 </State>
</TopologyEntry>
<TopologyEntry>
<ForPhones>
SILENCEPHONES
</ForPhones>
<State> 0 <PdfClass> 0 <Transition> 0 0.50 <Transition> 1 0.25 <Transition> 2 0.25 </State>
<State> 1 <PdfClass> 1 <Transition> 1 0.50 <Transition> 2 0.25 <Transition> 3 0.25 </State>
<State> 2 <PdfClass> 2 <Transition> 1 0.25 <Transition> 2 0.25 <Transition> 3 0.50 </State>
<State> 3 </State>
```

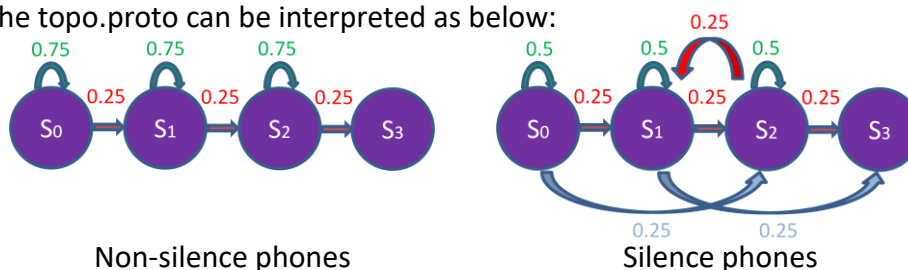
Accuracy: 74.96%

Runtime:

Preprocess	Extract feature	train	test
0	22secs	1min22secs	4secs

Observations and Hypothesis:

- (1) Most of time is spending on training. Therefore, when increasing the number of iterations, it is necessary to consider the time limitation of 5mins in total.
- (2) Number of Gaussians generated after training is too less.(only 4)
- (3) Optimal acoustic model weight and test beam might have minor effect.(not sure)
- (4) The topo.proto can be interpreted as below:



It is interesting to find that for silence phones, there is a probable transition from state2 to state1, which is unusual in HMM for speech recognition. Professor explains that this is because silence phones are already random, the states are allowed to go backward. I guess increasing the number of states might improve the accuracy.

## 2.Final Result and Discussions

```
### parameters that you can modify
numiters=12                                # Number of iterations of training
maxiterinc=300                             # Last iter to increase #Gauss on.
numgauss=600                               # Initial num-Gauss (must be more than #states=3*phones).
totgauss=10000                             # Target #Gaussians.
incgauss=$((totgauss-numgauss)/$maxiterinc) # per-iter increment for #Gauss
realign_iters="1 2 3 4 5 6 7 8 9 10 11 12";
scale_opts="--transition-scale=1.0 --acoustic-scale=0.1 --self-loop-scale=0.1"
###
```

```
### parameters that you can modify
opt_acwt=0.2
test_beam=10.0
###
```

Accuracy: 95.51%

Runtime:

Preprocess	Extract feature	train	test
0	23secs	3mins44secs	4secs

Discussions:

- (1) As expected, increasing the numbers of iterations and Gaussians significantly improves the accuracy because the base-line model doesn't complex enough. However, the time for training the model extends simultaneously. As the time limit exists, I can only set these values as high as possible.
- (2) Out of expectation, decreasing the acoustic model weight to 0.2 and test beam to 10.0 also improves approximately 5% of the accuracy. The best part is that the time for decoding remains pretty fast.
- (3) Although I believe that adding some states to topo .proto is helpful, I didn't modify it eventually since perhaps that is still not enough to pass over 99% accuracy and the runtime will enhance further.