

# Project6&7 report

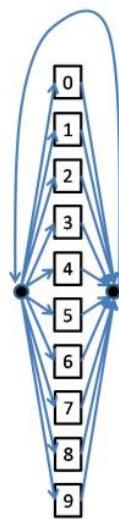
Yi Qiu  
14213830

## Project 6:

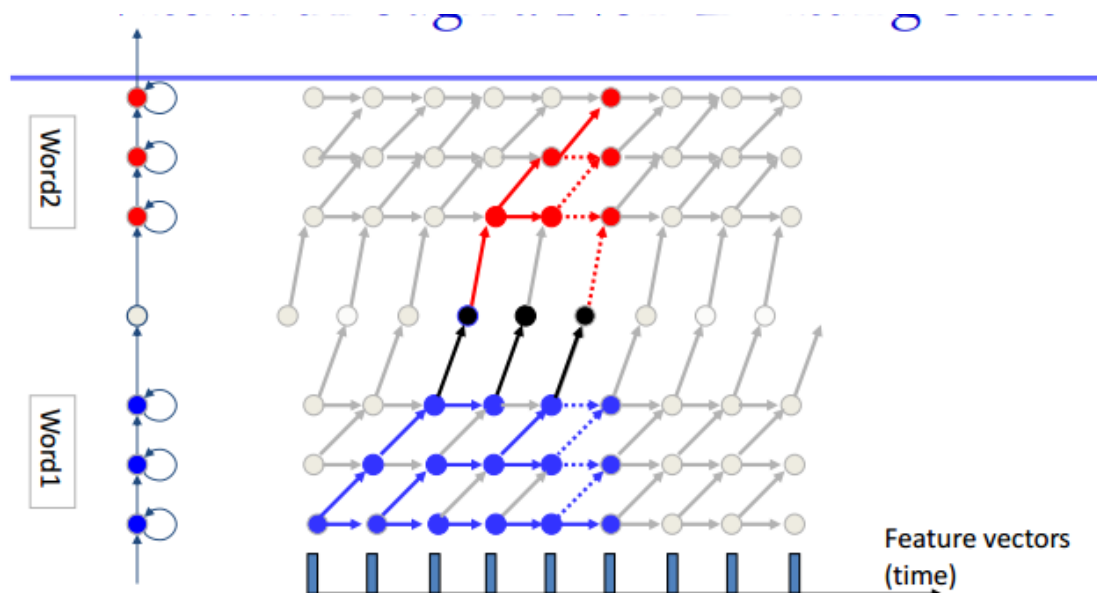
### 1: Design

In project 6, we have implemented a continuous speech recognition system based on isolated HMM model (with single Gaussian model). During the decoder process, we implemented the back pointer table and DTW to recognize the recorded words. And the decoder structure is as follows:

Back pointer structure:



DTW:



## 2: Test and accuracy

In this part, we do several tests and calculate their accuracy (these entire test data come from me and another group member). The result is as follows:

Original digit string: 911385

Result	Digit accuracy	String Accuracy
Result.1 515165	2	0

Original digit string: 826414052002

Result	Digit accuracy	String Accuracy
Result.1 606116656666	3	0

Original digit string: 8212176342

Result	Digit accuracy	String Accuracy
Result.1 6212166662	6	0

Original digit string: 7343332190377

Result	Digit accuracy	String Accuracy
Result.1 7616766166367	4	0

Original digit string: 2212

Result	Digit accuracy	String Accuracy
Result.1 2212	4	1

Original digit string: 123456

Result	Digit accuracy	String Accuracy
Result.1 126116	3	0

Original digit string: 6890372344

Result	Digit accuracy	String Accuracy
Result.1 6650660301	3	0

Original digit string: 72184347924

Result	Digit accuracy	String Accuracy
Result.1 72166767234	5	0

Original digit string: 5555

Result	Digit accuracy	String Accuracy
Result.1 5555	4	1

Original digit string: 37274921

Result	Digit accuracy	String Accuracy
Result.1 37277501	5	0

From these results, we can see conclude:

Sentence accuracy:  $2/10 = 20\%$

Digit accuracy:  $39/94 = 42\%$

From the result we can see that the digit accuracy is unacceptable since in our recognition system, silence model is not considered.

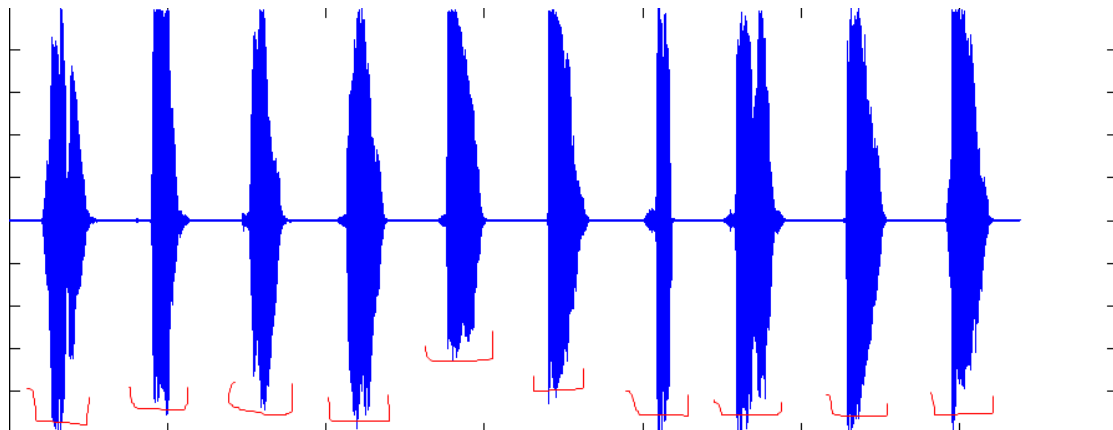
## Project7

### Problem1

#### 1: Design

In project 7, we have implemented a continuous speech recognition system based on continuous HMM model. During the decoder process, we implemented the back pointer table and DTW to recognize the recorded words. And the structure is pretty much like the structure in project 6.

In fact, the basic method would be same as the project6. We also use the same DTW process, but we need to change our training data. First, we would use the isolate model to detect the continuous recording, so that we could at least detect how many numbers we said in the recording. The image of detail is below.



#### 2: Test and Accuracy

In this part, we do several tests and calculate their accuracy:

Using above method, we test the program again, the test data would be the same as project6.

Here is our result.

Original digit string: 911385

Result	Digit accuracy	String Accuracy	Digit accuracy: 4.2 Percentage = 70%  String accuracy: 0.2
Result.1 911385	5	1	
Result.2 9113897	5	0	
Result.3 9113897	5	0	
Result.4 27	0	0	
Result.5 9113853	6	0	

Original digit string: 826414052002

Result	Digit accuracy	String Accuracy	Digit accuracy: 10.6 Percentage = 88.3%  String accuracy: 0
Result.1 6826414092002	11	0	
Result.2 6266414092002	11	0	
Result.3 6266414092002	11	0	
Result.4 6866414092002	10	0	
Result.5 6226414092002	10	0	

Original digit string: 8212176342

Result	Digit accuracy	String Accuracy	Digit accuracy: 7.2 Percentage = 72%  String accuracy: 0
Result.1 8212176	7	0	
Result.2 8212176	7	0	
Result.3 8212176	7	0	
Result.4 82121763525	8	0	
Result.5 8212176	7	0	

Original digit string: 7343332190377

Result	Digit accuracy	String Accuracy	Digit accuracy: 2.2 Percentage = 16.9%  String accuracy: 0
Result.1 7343332190377	2	0	
Result.2 7343332190377	1	0	
Result.3 7343332190377	4	0	
Result.4 7343332190377	2	0	
Result.5 7343332190377	2	0	

Original digit string: 2212

Result	Digit accuracy	String Accuracy	Digit accuracy: 4 Percentage = 100%  String accuracy: 5
Result.1 2212	4	1	
Result.2 2212	4	1	
Result.3 2212	4	1	
Result.4 2212	4	1	
Result.5 2212	4	1	

Original digit string: 123456

Result	Digit accuracy	String Accuracy	Digit accuracy: 6 Percentage = 100%  String accuracy: 5
Result.1 123456	6	1	
Result.2 123456	6	1	
Result.3 123456	6	1	
Result.4 123456	6	1	
Result.5 123456	6	1	

Original digit string: 5555

Result	Digit accuracy	String Accuracy	Digit accuracy: 4 Percentage = 100%  String accuracy: 5
Result.1 5555	4	1	
Result.2 5555	4	1	
Result.3 5555	4	1	
Result.4 5555	4	1	
Result.5 5555	4	1	

Original digit string: 37274921

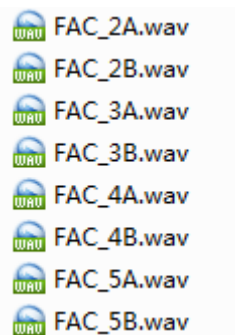
Result	Digit accuracy	String Accuracy	Digit accuracy: 7 Percentage = 87.5%  String accuracy: 0
Result.1 372774521	7	0	
Result.2 372774521	7	0	
Result.3 372774521	7	0	
Result.4 372774521	7	0	
Result.5 372774521	7	0	

It is obvious that the accuracy is better than before, which means the continuous model is work.

## Problem2

This problem, we are still on the way. So we only report what we have done so far.

First, we would extract those isolated recording, just like the below.



These recordings would be trained first, so that we can ensure our program can adapted these recordings.

And then, we would train those recordings that their number of words is detected by program is same as the original recording. Finally, we train the rest of recordings.

## Feedback

Overall, this course is pretty cool and by doing several projects, I have improved my programming skills since I barely programed before.

But since given the fact that we students have different academic backgrounds, maybe handouts of projects can have more details in programming.