

# DSP HW3 Report

B05901172 黃禹靖

- Environment:

Linux ( bash on window )

C++11

Machine type: i686-m64

- Compile:

Generate ZhuYin-Big5.map: make map

Compiling mydisambig: make

clean: make clean

- Execute:

run all 10 text file: make run

or ./mydisambig -text [corpus.txt] -map ZhuYin-Big5.map -lm bigram.lm -order 2  
>[outputfile]

- Discussion:

The size of a Chinese word in Big5 encoding is 2 bytes. And the size of char in C++ is 1 byte. So for every Chinese word, I use two char to represent it and stream in or out two char every single time.

The procedure of mydisambig completely follows Viterbi algorithm refer in class. However, the result of mydisambig is little different from result of disambig of srilm, where the differences can only be at first or last words of sentences. That is interesting, but I can't figure out the reason.