

Digital Speech Processing Homework #2-2


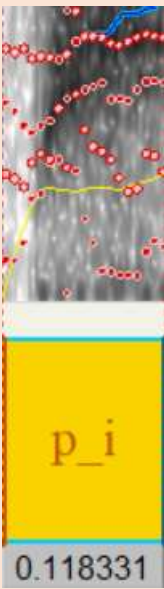

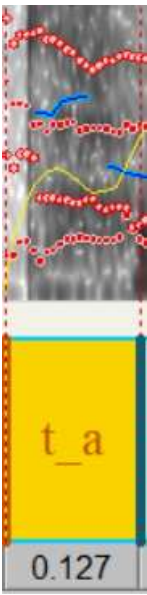
Speech Analysis

Student ID: R11945009

Name: 周知頤

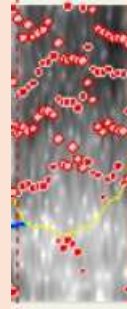
Part 2:

Phonetic Class

Plosive/Stop	p(ㄆ)	 <p>0.117426</p>	 <p>0.118331</p>
Plosive/Stop	t(ㄊ)	 <p>t_o</p>	 <p>t_a</p> <p>0.127</p>

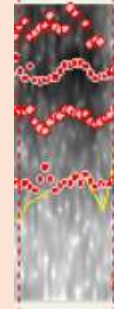
Fricatives

sh(尸)



sh_a

0.142

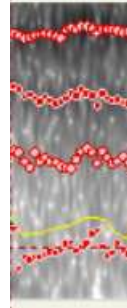


sh_e

0.120

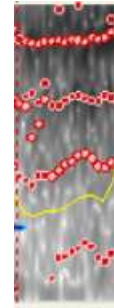
Fricatives

shi(ㄣ)



shi_i

0.153



shi_i

0.097

Affricates

ch(ㄷ)



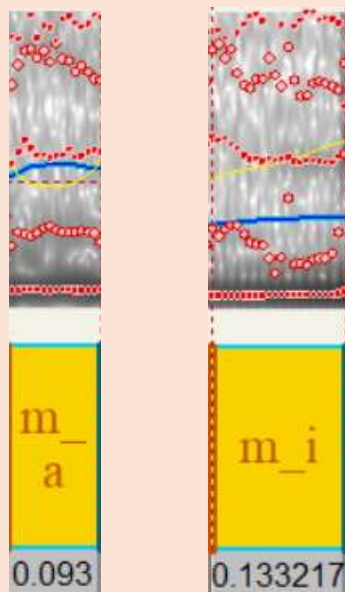
Affricates

ji(ㅈ)



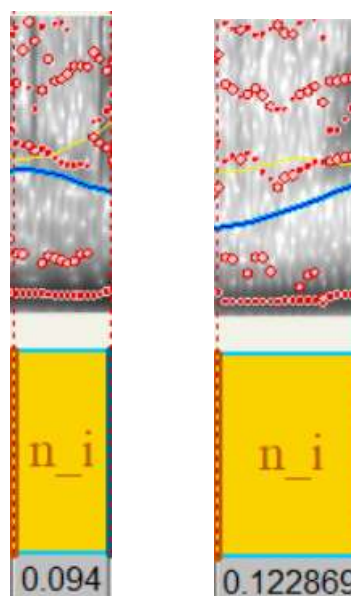
Nasals

m(ㄇ)



Nasals

n(ㄋ)



Part 3:

1. What are the consistencies of the spectrogram in each phonetic class? (Plosive, Fricative, Affricate, Nasal)?

Plosive 的部分可以觀察到都是由淺轉深。突然有一股能量的爆發，造成發聲的開始。Fricative 的部分有點像噪音，感覺是因為湍急的氣流而造成頻率很混亂。Affricate 感覺是介於 plosive 和 fricative 中間，因為他同時擁有兩者的特徵。Nasal 則是在高頻的地方沒有太多的振幅。

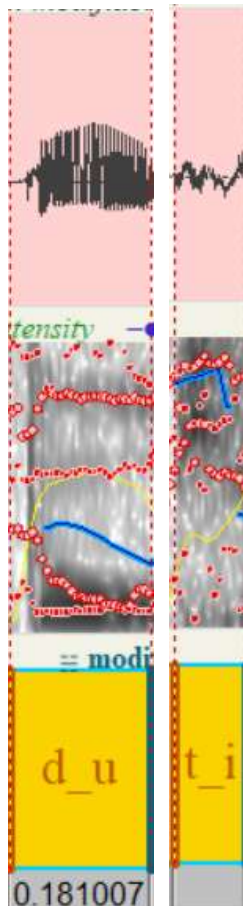
2. Is the boundary between neighboring initial and final clear? What is the benefit of using “right-context dependent” initial model (ex: sh_a) instead of pure initial model (ex: sh) to model initials? Please explain with examples.

有時候界限其實很模糊。使用 sh_a 而不是只有 sh 的優點是，在標注像是

「傻」的音時，因為讀音很難將 sh 與 a 分開，導致分割時界限可能會非常模糊。而使用 sh_a 與 a 可以增加分割的誤差值，更方便我們分割。

3. What are the differences when pronouncing ㄅ & ㄆ? How can you tell the differences in spectrogram for ㄅ & ㄆ? (You may also want to compare ㄆ & ㄇ, ㄆ & ㄇ respectively)

ㄅ&ㄆ&ㄇ 發音需要用到聲帶的震動，而ㄆ&ㄇ&ㄇ比較像是用氣音的方式。Spectrogram 上可以看到需要用到聲帶的震動的發音振幅較強烈。



4. Take a look at the spectrogram of finals. Is there any simple rules to discriminate initials from finals provided only spectrogram?

我認為 finals 大部分都需要聲帶的震動，而配合 initials 大部分振幅較小。其實很容易就可以先大概區分 initials 和 finals。若是整個粘在一起的話，才需要再花點時間聽聽看聲音。

BONUS:

我認為該作品為《竹聲桃花》。首先觀察 spectrogram，第三個音擁有 Plosive 的特徵而第二個音的 initial 有點拉長音且振幅適中所以認為是 Fricative initial。結合這兩個特徵篩選出川端康成的三個作品，分別為《女性開眼》、《晚霞少女》和《竹聲桃花》。接下來看到第一個音，該音的 formants 結構並不像是黏在一起的音，反而是稍微分開。「女」和「晚」發音較沒有「竹」來的好區分，因此將《女性開眼》、《晚霞少女》淘汰，留下《竹聲桃花》。