Speech technology, formants and phonemes extraction

Dekel Viner

October 2016

1 Exercise 1

1.1

268Hz

1.2

1. "Hi" is pronounced the loudest
2. "I'm" has the highest pitch
3. 3529
4. at $\text{Time}_s 1.623492 the first 4 formants are$:
   $F1 : 654.801175 Hz$
   $F2 : 1837.929980 Hz$
   $F3 : 2488.526478 Hz$
   $F4 : 3746.255054 Hz$

2 Exercise 2

2.1

At these frequencies we can generally only hear the first formants of the sounds, so a lot of information that we derive from the second,third, enc. formants is lost and we cannot use it to understand the speech.

2.2

The quality of the sound is certainly good enough to understand. between 300 and 3000Hz we still have most of the information, for many sound we will still

have the fundamental frequency and the second and third formants which are enough provide the largest amount of information. For certain sound we will loose the fundamental frequency but generally we can still interpret the sound from the general context and the rest of the formants that are available as they still provide a substantial amount of information.

3 Exercise 3

4 Exercise 4

4.1

first formant: 469.1369158764888 Hz (mean F1 in SELECTION)
$f_1 = c/4L$
$L = c/(f_1 * 4)$
$345/4F_1 = L = 0.184 = 18.4cm$

4.2

/a/ and /i/ make chambers in the vocal tract when produced that would modify the formant frequencies. /i/ closes the dental part of the mouth completely and /a/ keeps it wide open. /3r/ on the other hand keeps the vocal tract open at the same the same consistency all along the way (no chambers), so we can use it to calculate the length of the entire vocal tract with its formant frequencies.

4.3

Lip rounding would allow more air to flow back into our mouth from the high pressure outside of the mouth into the low pressure in side of our mother when air goes through it. This would increase the frequency of our voice. With the formula that we have been using he have regarded the vocal tract to be a closed tube lip rounding would change that.

5 Exercise 5

6 Exercise 6

6.1

When looking at the plot it might be kind of hard to distinguish that there are two different type of l, but there is a lot of variability on the second formant

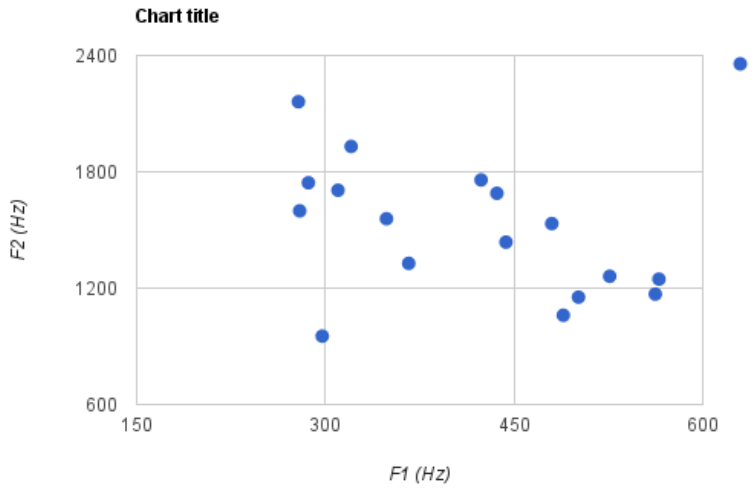


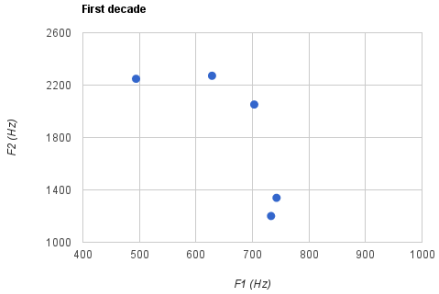
Figure 1: First and second formants of the l sounds

(ranging between 952-2161) we can distinguish two groups averaging around 1200 and 1800 Hz on the second formant. We can find a number of words in albanian that are minimal pairs with regards to the l sounds for example, there are two words that sound like "vi-el". One has first and second formants (373.704300, 1727.482491) and the other (575.327102, 1283.146146), another example are two words that sound like "fail" one of them has first and second formants(550.082147, 1236.625600) and the other (406.750276,1832.355172). Clearly there are two different l types in albanian. In English however we cannot find such minimal pairs.

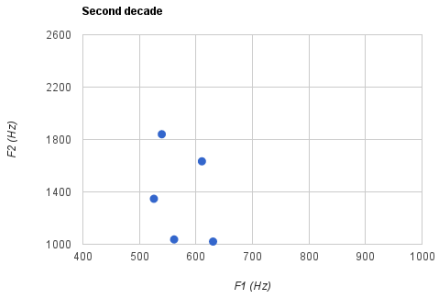
6.2 Exercise 7

6.3 Exercise 8

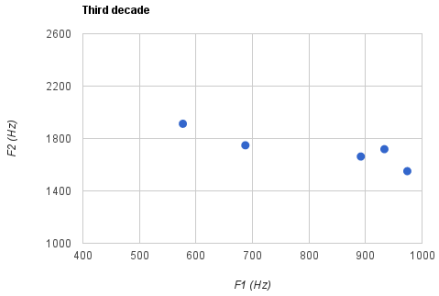
In the first decade her voice was higher and difference between the /æ/ and /ɛ/ sounds were more noticeable, on the third decade her voice became lower and on the final decade it was became harder to spot the /æ/ sound, they changed to /ɛ/, as perhaps it requires more effort to produce the /æ/ sound, and with old age she replaced it with /ɛ/.


First decade
2600
2200
1800
1400
1000
F2 (Hz)
400
500
600
700
800
900
1000
F1 (Hz)

(a)


Second decade
2600
2200
1800
1400
1000
F2 (Hz)
400
500
600
700
800
900
1000
F1 (Hz)

(b)


Third decade
2600
2200
1800
1400
1000
F2 (Hz)
400
500
600
700
800
900
1000
F1 (Hz)

(c)