Speech Processing Project 1 Writeup

Matt Ruffner

February 15, 2019

1

Choosing a frame size of 512 simplified things in that I was able to use nFFT equal to the frame size for the majority of the project. It also provided a decent (50ms) temporal resolution. 75% overlap was used in the spectrogram to obtain artificially high temporal resolution and decent appearance (compared to a Praat spectrogram, for example).

2

Energy and zero crossing rate seem to be an excellent feature for representing the voiced/unvoiced status of a phoneme. As shown in the plots, the zero crossing rate is substantially higher for unvoiced phonemes than for voiced phonemes. Conversely, the signal energy is much higher for voiced phonemes than for unvoiced phonemes.

5

There is a clear peak visible for the two voiced phonemes of the letters "u" (υ) and "n" (n). The f_0 of the υ phoneme is 1582 Hz, while the peak in the cepstrum corresponding to the f_0 of the n phoneme is at 1152 Hz.

6

A cutoff of 50 was used because this was the position within the cepstra which accurately separated the vocal tract filter information from the excitation information for both of the voiced phonemes.

7

I think the order 14 gives the most accurate information regarding the vocal tract filter. The 40th order filter provides an over-fitted model and the main structure of the filter is not as easily interpreted.

9

The formant tracks and fundamental frequency contour of the `sun.wav` audio file is shown in Fig. 1. The red dots show the formant tracks and the blue line is the f_0 contour.

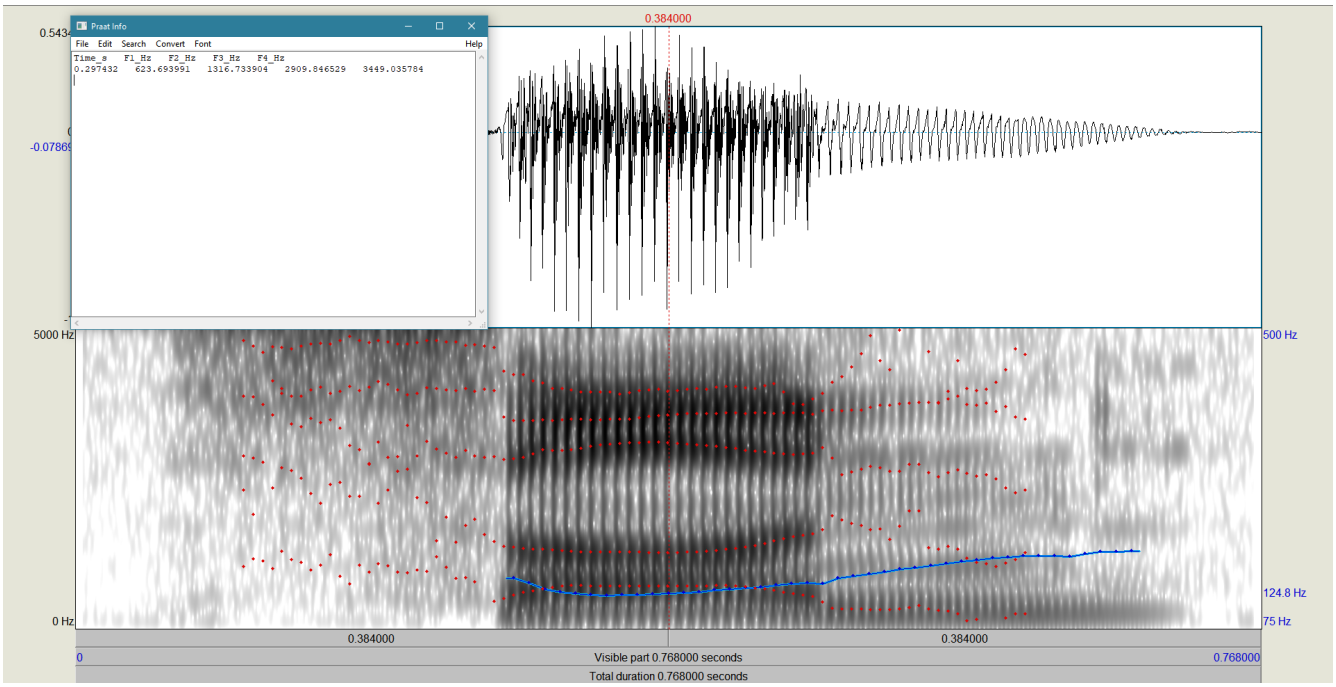

Figure 1: Spectrogram of `sun.wav` with f_0 contour and formant tracks.

10

In this section, three tables are presented which show the formant frequencies as they are estimated from the various plots that were produced as a result of the exercises in this project. The formants in Table 1 are for the s phoneme, the formants in Table 2 are for the υ phoneme, and the formants in Table 3 are for the n phoneme.

	DFT	Lift. Spec.	LPC (4)	LPC (14)	LPC (40)	Praat
F_1	429 Hz	390 Hz	-	1094 Hz	781 Hz	764 Hz
F_2	1035 Hz	605 Hz	-	2715 Hz	1133 Hz	2190 Hz
F_3	3008 Hz	1035 Hz	-	-	2031 Hz	3025 Hz

Table 1: Formant Frequency comparison for the s phoneme.

	DFT	Lift. Spec.	LPC (4)	LPC (14)	LPC (40)	Praat	Literature[1]
F_1	625 Hz	703 Hz	1465 Hz	1348 Hz	486 Hz	704 Hz	813 Hz
F_2	1250 Hz	1211 Hz	-	2480 Hz	1289 Hz	1260 Hz	1323 Hz
F_3	3164 Hz	3086 Hz	-	-	2500 Hz	3078 Hz	-

Table 2: Formant Frequency comparison for the υ phoneme.

	DFT	Lift. Spec.	LPC (4)	LPC (14)	LPC (40)	Praat
F_1	156 Hz	137 Hz	391 Hz	468 Hz	352 Hz	309 Hz
F_2	957 Hz	1016 Hz	-	2070 Hz	684 Hz	1480 Hz
F_3	2070 Hz	1641 Hz	-	4180 Hz	2070 Hz	2612 Hz

Table 3: Formant Frequency comparison for the n phoneme.

The results were somewhat varied, with the liftered cepstrum and 40th order LPC matching praat for on several occasions. In some instances, the order of the LPC was too low in order to predict a third or even second formant frequency. The somewhat varied results like this could be in part due to the lack of overlapping windows, or poor interpretation of which peaks correspond to formant frequncies.

Varying the order of the LPC algorithm greatly varied the frequency response of the resulting filter, in some instances, I believe there was an overfitting of sorts which led to the misrepresentation of formant frequencies. This is in contrast to a low order LPC, like 4, which provides little to no information about the formant locations (a single formant, at most).

On average, the liftered cepstrum most often matched the Praat f_0, including the case of the vowel phoneme, υ. The most accurate predictor that was done in this project seems to the the liftered cepstrum. This was a very fun project, well rounded and cohesive, each task that was a assigned had a clear motivation and required explanations were warranted.

[1] http://ec-concord.ied.edu.hk/phonetics_and_phonology/wordpress/learning_website/chapter_2_vowels_new.htm