I analysed the sound with quite a big window size, around 3001, and a bigger FFT size. What I can see, is that the sound seems to have a very important harmonic structure (the parallel lines in the spectrum). The lowest frequency, which with very high probability will be the fundamental, ranges between approximately 120 and 260 Hz. The number of harmonics seems to depend on the phonemes: some of them, specially vowels, have defined harmonics up until 10kHz (thus maybe having up to 100 harmonic in those cases), while others (for example the s), are practically pure stochastic.

Part 1.2:

* window type: blackman
Blackman has 6 bins main lobe and smaller side lobes, which proves to be adequate to detect the f0 and the harmonics well.

* window size: 1501
I found that this size worked good, when dealing with the stft. Lower values dont recognize enough harmonic information, and higher works well but might be a little worse on attacks.

* FFT Size: 4096
This was a size that gave enough resolution, and also 2048 had produced some artifacts in the harmonic detection.

* minimum f0: 120
  I could see in the spectrogram that minimum f0 was more than 120Hz. The f0 detection is working pretty well

* maximum f0: 260
  I could see in the spectrogram that maximum f0 was less than 260Hz. The f0 detection is working pretty well

* error threshold in f0 detection: 5
  Smaller thresholds will start to not recognise properly the fundamental

* maximum number harmonics: 60
  Even though I think the ideal should be around 100 harmonics, I think this compresses a lot of data while keeping

similar quality at the output.

* stochastic decimation factor: 0.3
  We achieve a good rate of compression, while still not changing the output quality very much.