Questions:

Preliminary Question: “Given a sampling rate of 8000Hz, compute the length in samples of a window used for narrowband analysis, i.e., a window of duration 25ms, and call that length Nwn. Compute the length in samples of a window used in wideband analysis, i.e., a window of duration 3ms, and call that Nww.”

The length of the window (in samples) can be calculated as the following:

Question:

1. Try to determine from your plots how you can use Q1 and Q2 to segment the sentence into

silence, vowels and consonants. Compare with the plot of data.

Just looking at the amplitudes, Qa (/a/ sound) had the highest short-time energy, and Qsh (/sh/ sound) had slightly lower energy. Theoretically, a segment of silence should look like an energy / amplitude very close to zero. Vowels should have higher short-time energy levels than consonants, and so a threshold in energy can be applied (assuming that the voice level has been first normalized, of course. Obviously, we wouldn’t want someone yelling into a microphone to always be recognized as vowels-only).

2. Compare the plots of Q3 and data and see if you can set a threshold on Q3 that can be used to discriminate vowels from noise/silence/fricatives. Use the plots of Qa and Qsh to set those thresholds.

ZCR (Zero-crossing-rate) is known to have higher amplitudes for unvoiced sounds than voiced sounds. This is because unvoiced sounds are noisier in the higher frequencies. Hence, ZCR can be used as a threshold to differentiate between voiced and unvoiced sounds. (Noise / Silence respectively). In our plots, you can observe that the ZCR was higher for Qsh (/sh/ sound) than for Qa (/a/). Any peaks past 1.5 for /a/ should be used on the ZCR plot, and for the /sh/ plot peaks past 140.

TODO: Check this

3. Feel free to use different window types and window lengths and see what effect they have on your results.

If a 5 ms window is used, this changes the window length to be 40 samples Using this window instead of the 200 sample window of .025 seconds, we see higher energy and amplitude for /a/, and a higher ZCR for /sh/. A smaller window creates fewer smooth contours than the 200-sample window. Please reference Figures 7 and Figures 8.

4. Note that Ra is the autocorrelation of a, and Rsh is the autocorrelation of sh. Compare Ra with AMDFa and Rsh with AMDFsh.

TODO: Do this

5. For each signal (“female a” and “male a”), at approximately what window size, Nw, are

harmonics no longer resolved? What is the spectral resolution (DFT-bin spacing) in Hz

corresponding to these window lengths?

For male and female signals, the window lengths when harmonics are no longer resolved are found just by plotting different window sizes in matlab. This ended up corresponding to L = 90 samples and L = 50 samples. In terms of spectral resolution, it resolved to the following Hz:

Male:

Female:

TODO: Make sure that these L values are correct. It’s possible the datasets changed.

Plots:

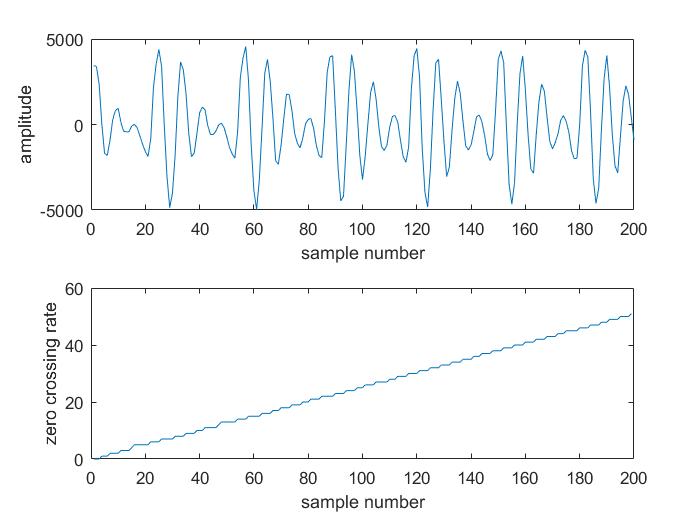


Figure 1: Q1.

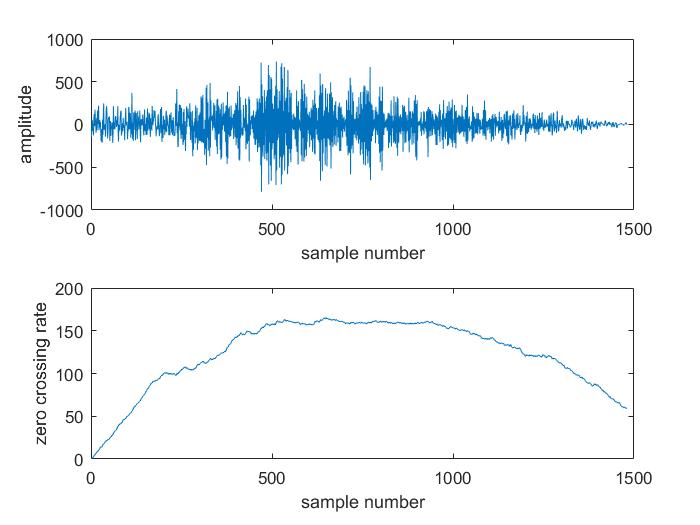


Figure 2: Q2.

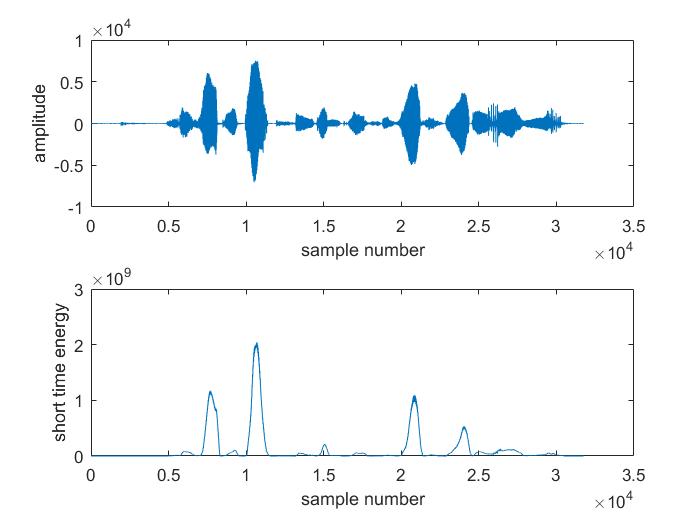


Figure 3: Q3.

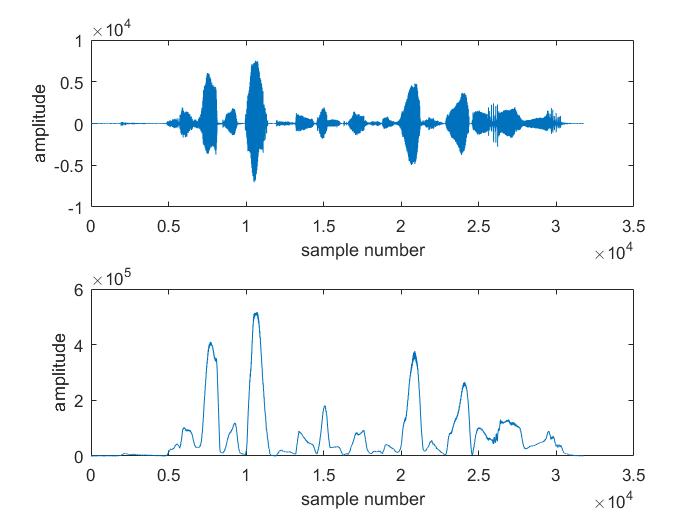


Figure 4: Qa with Nwn of 200 samples.

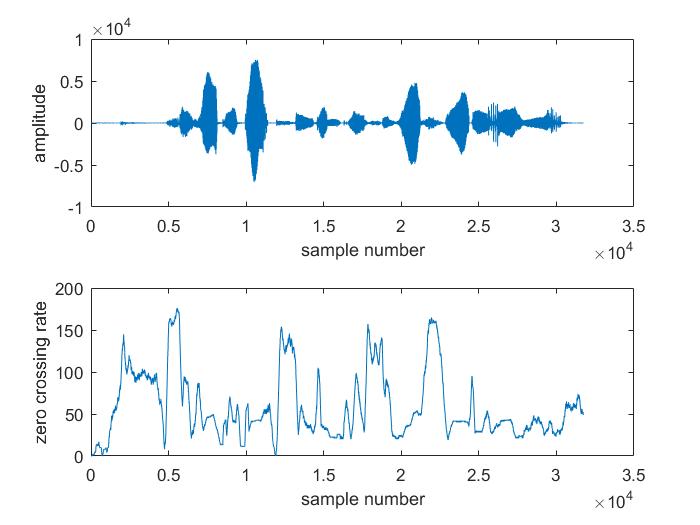


Figure 5: Qsh with Nwn of 200 samples

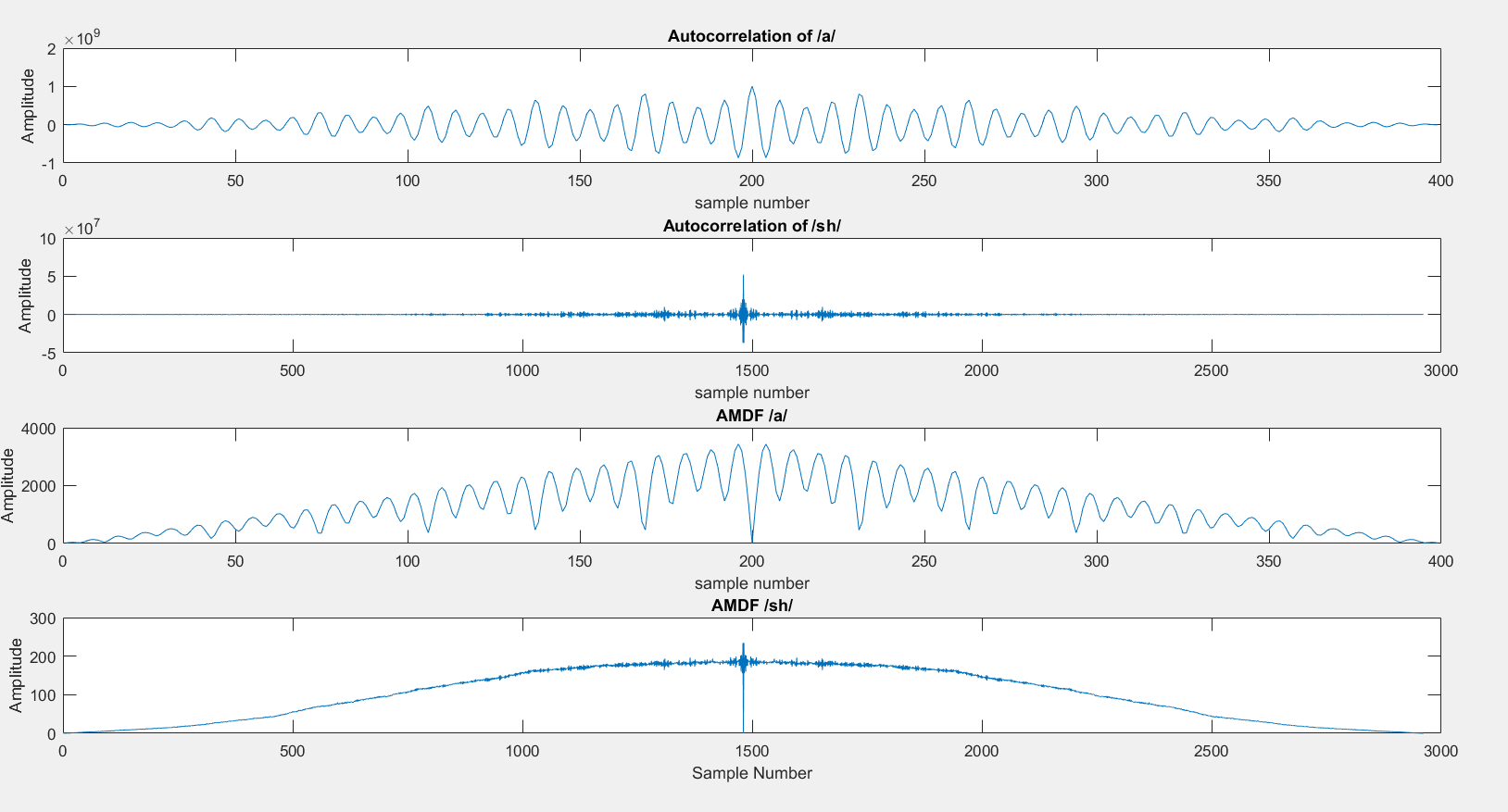


Figure 6: Autocorrelations and AMDFs of /a/ and /sh/

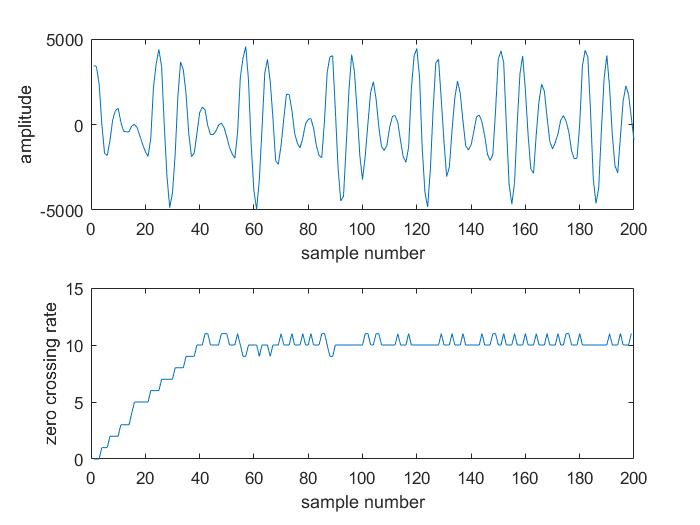


Figure 7: Qa with Nwn of 40 samples

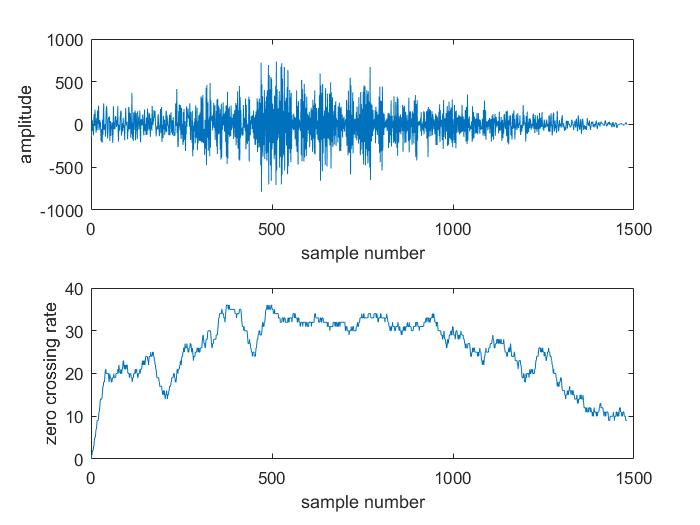


Figure 8: Qsh with Nwn of 40 samples

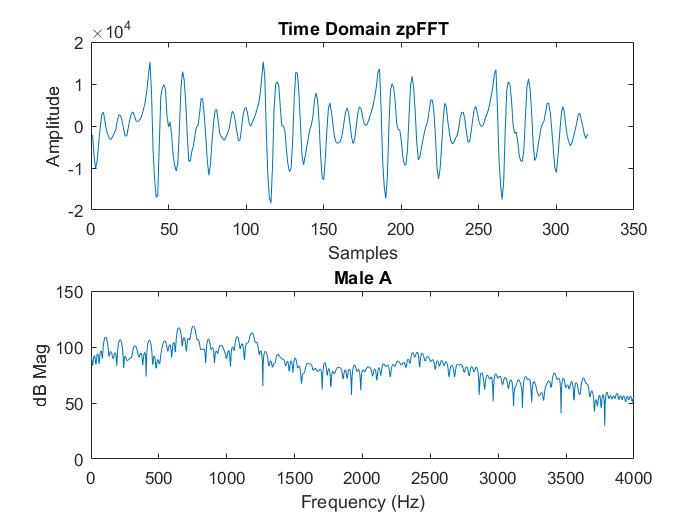


Figure 9: Male A. Evidently there are some harmonics that are very visible in the lower spectrum of this graph ( less than 1500 Hz)

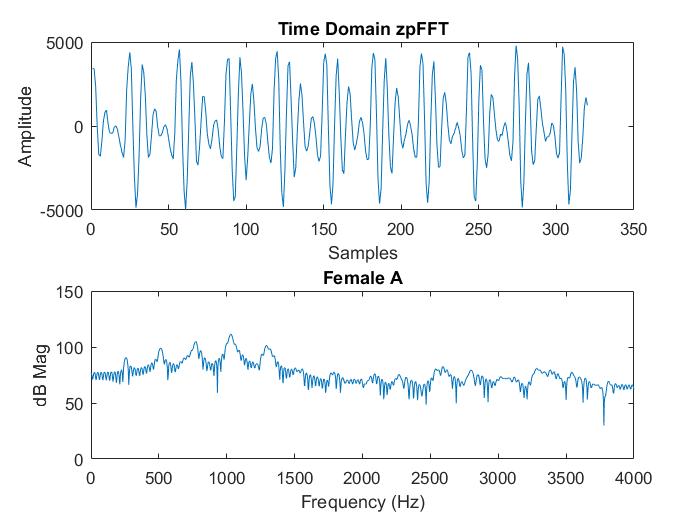


Figure 10: Female A. Evidently harmonics exist at about every 250 Hz.

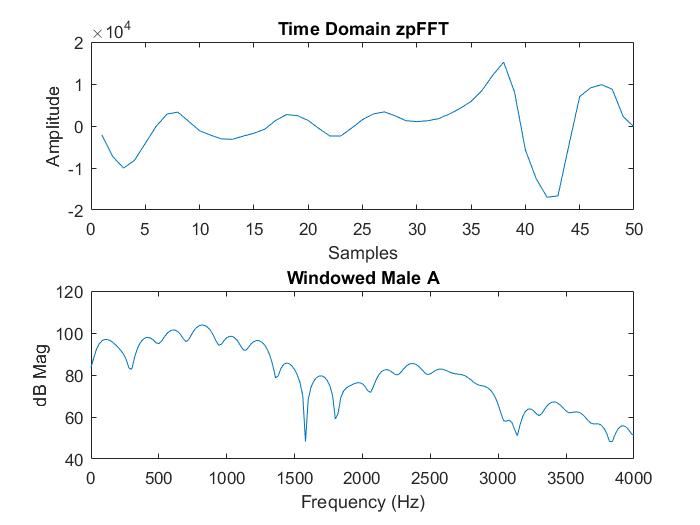


Figure 11: Windowed Male A

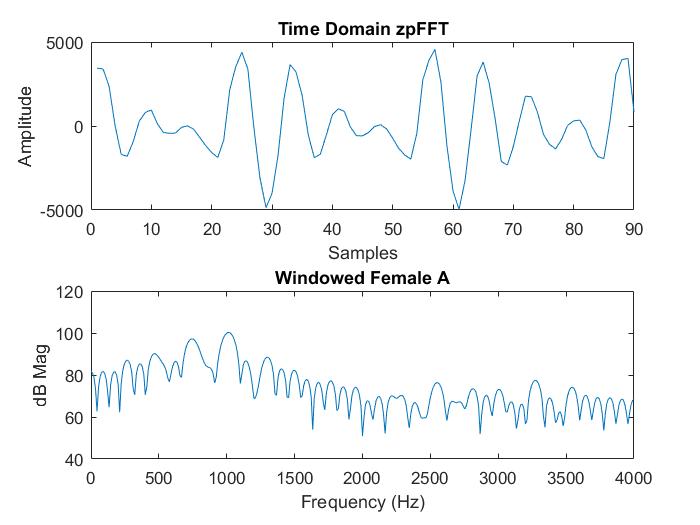


Figure 12: Windowed Female A