EE 513 HW 7

Matt Ruffner

April 10, 2017

1 Problem 7.1

An example spectrogram, generated for a 'yes' recording, can be seen in Fig. 1. An example of a 'no' recording can be seen in Fig. 2.

Determining if formant frequencies are present and if there is more spectral distribution of energy, as might be present in unvoiced speech will be used as criteria for the beginning and ends of the recording

The script that iterates over input files in the `yesnotest/` directory and plots their respective spectrograms is titled `loopSpectro.m` in the HW submission

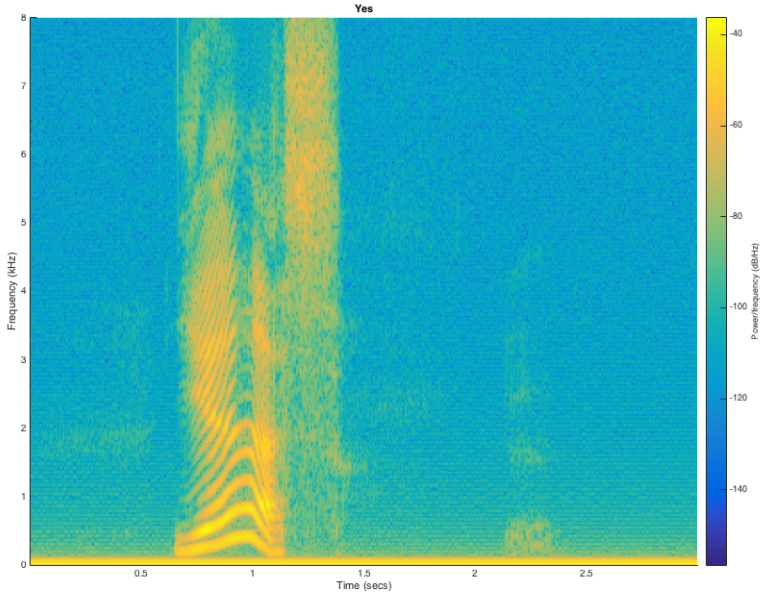

Figure 1: Spectrogram of a recorded 'yes'.

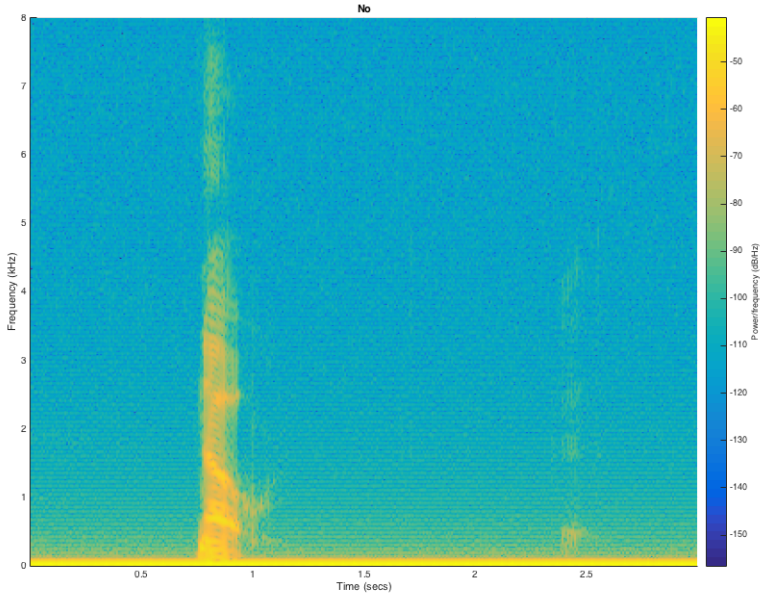

Figure 2: Spectrogram of a recorded 'no'.

2 Problem 7.2

The function that was written to extract the spoken voice portion of the provided clips is titled `extractUtterance.m` within the HW submission.

This function takes in the sample data and sample rate, preforms some signal conditioning and returns a subset of the provided data. The envelope of the signal was determined after having removed low frequency standing waves from the audio. The envelope was smoothed with a final low pass filter before the index at which the maximum signal value occurred at was found.

The indices at which the final signal data is to be extracted from are determined by searching forward and backward from this maximum index, waiting for the signal to fall below a 50th of the maximum signal value for 500 and 1000 samples, respectively.

Examples of good trimming performance, both before and after, can be seen in Figs. 3 and 4, respectively. Examples of sub-par performance, before and after trimming, can be seen in Figs. 5 and 6, respectively.

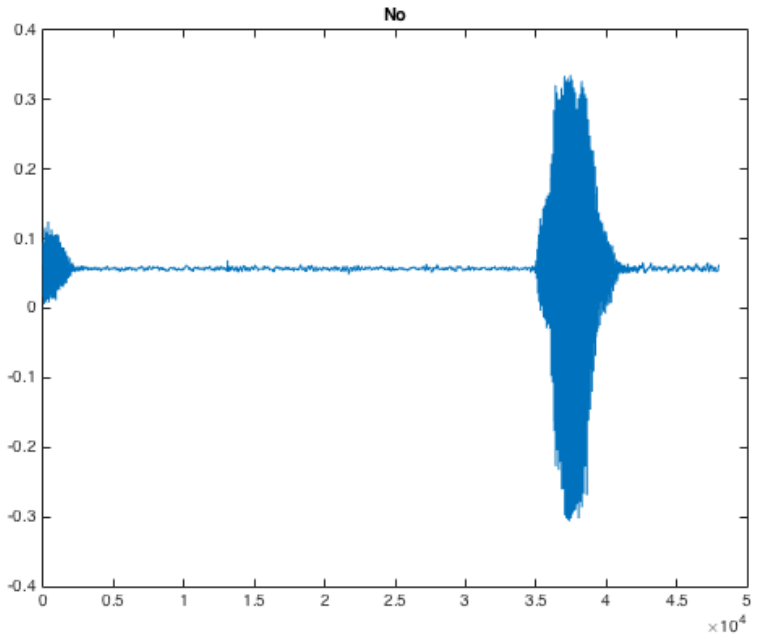

Figure 3: Original signal before trimming; correctly trimmed scenario.

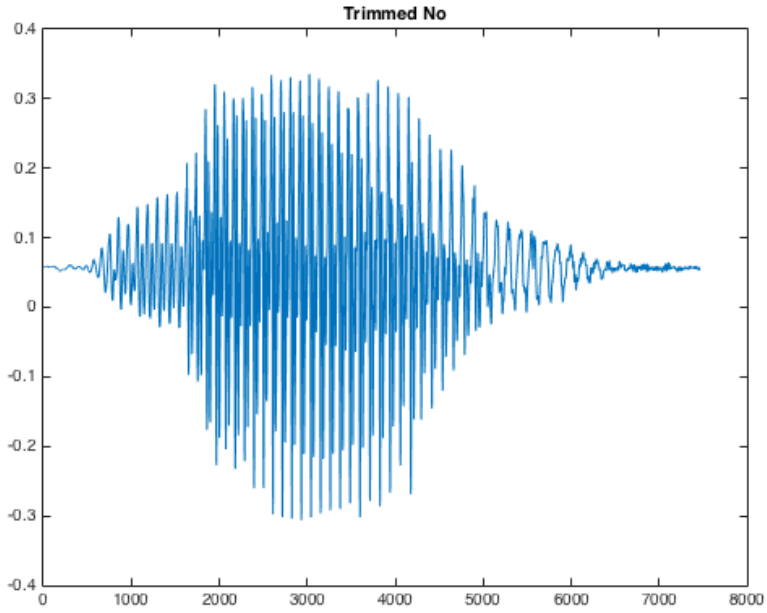

Figure 4: The correctly trimmed signal, not getting confused by beginning blip.

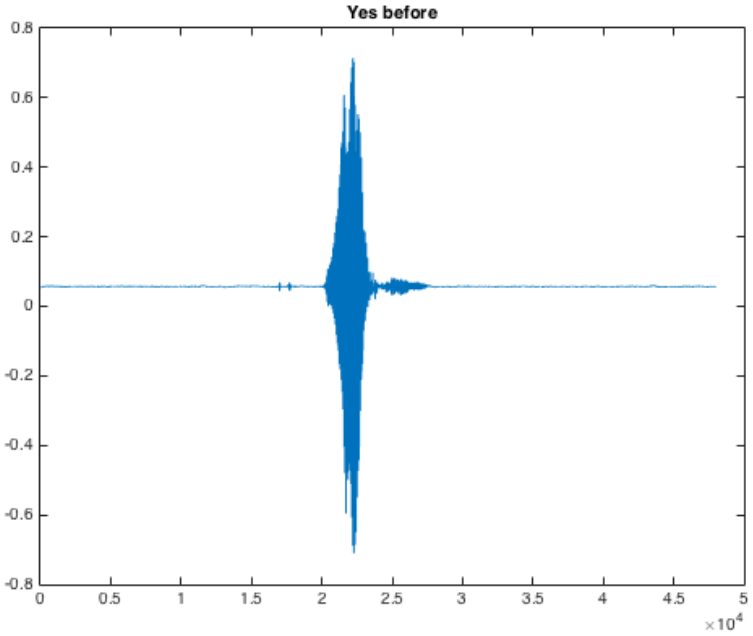

Figure 5: Inadequately trimmed scenario, before trimming.

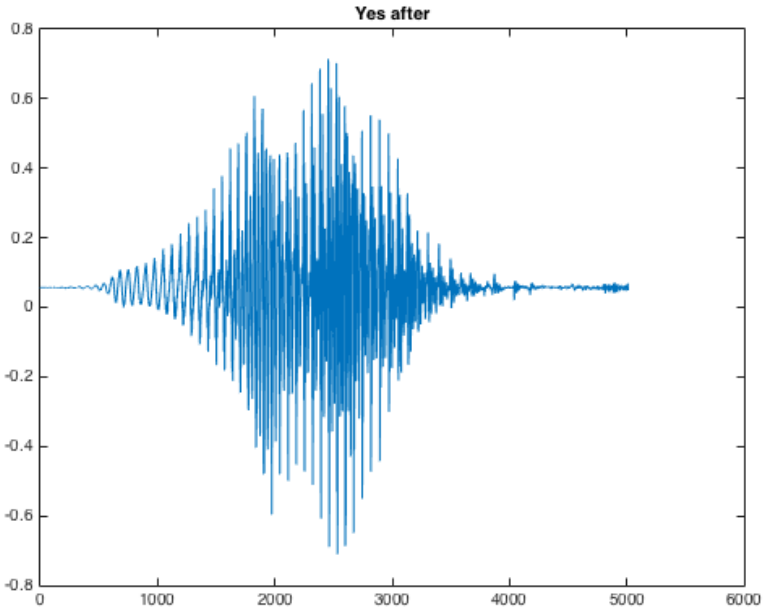

Figure 6: After trimming, a small portion of the unvoiced speech was cut out

3 Problem 7.3

For this problem, the function `estFormants.m` was created to extract the formant frequencies for the first and second half of a recorded segment of speech. The ratios of the second to fourth formant frequencies, in dB, were compared for the first half of the word to the second half. This metric, over many test cases, produced the feature plot seen in Fig. 7.

The script is tested in the test script `testEstForm.m`.

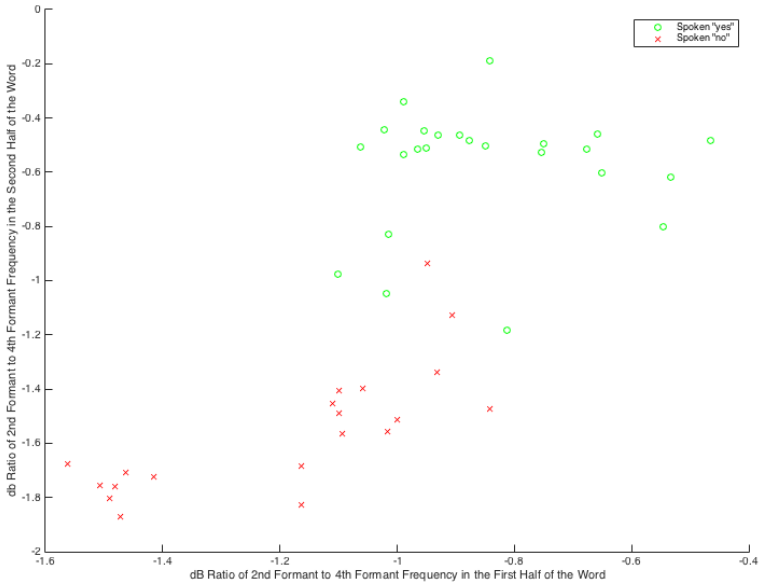

Figure 7: The resulting feature plot from running `testEstForm.m`.