**University of Sheffield**

COM3502

Speech Processing Report



**Professor Roger K. Moore**

Samuel Donnan

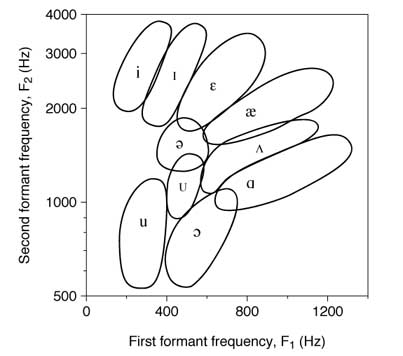
Muhammad Haroon Chishty

(50:50)

# Principle used to estimate parameters

## Formants

To begin, the file “Example 2-3 frequency measurement.pd” was used in experiments conducted to make sure the frequency reading was accurate. Human voices producing different vowels were used as samples. The output values were compared to those values produced from other means of frequency measurement such as the app “SpectrumView” available for iOS devices, developed by Oxford Wave Research Ltd. After recording the frequency measurement values, they were compared to Figures 1 and 2. A few minor tweaks were then implemented and it was believed that the readings were then more accurate.

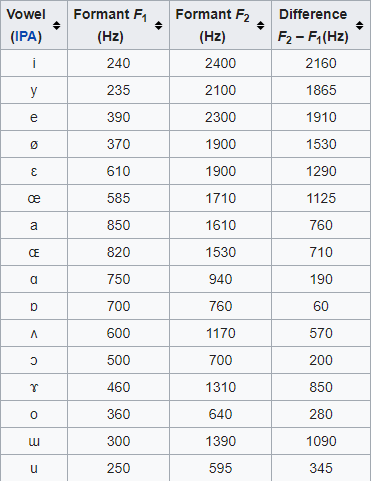


**Figure 1**

**The National Centre for Voice and Speech (America)**

<http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/filter.html>

A chart displaying F1 and F2 frequencies and tongue placement when producing different vowels



A table showing different values for frequencies in Hz given from producing different vowels

**Figure 2**

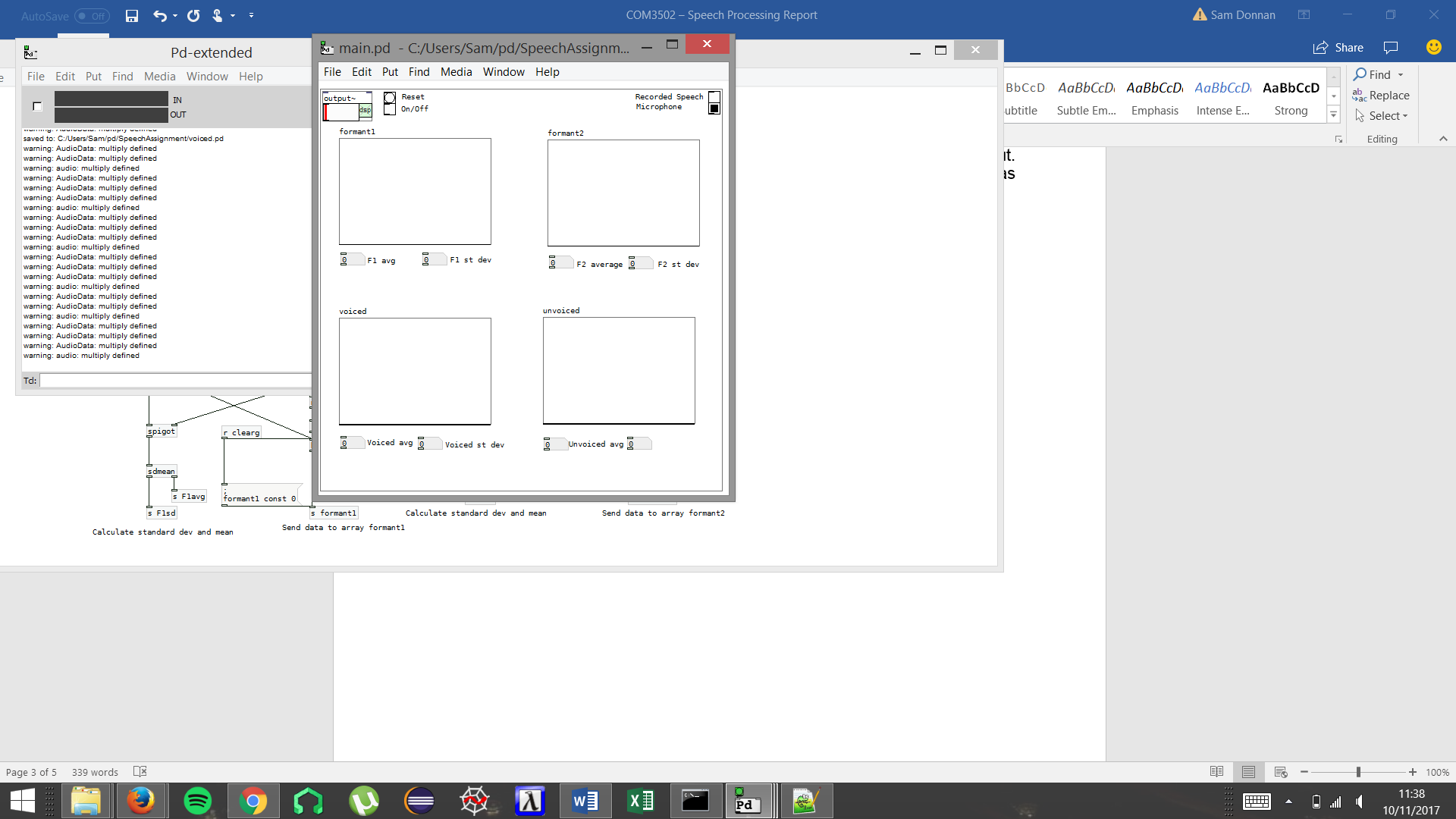
**Wikipedia - Formant**

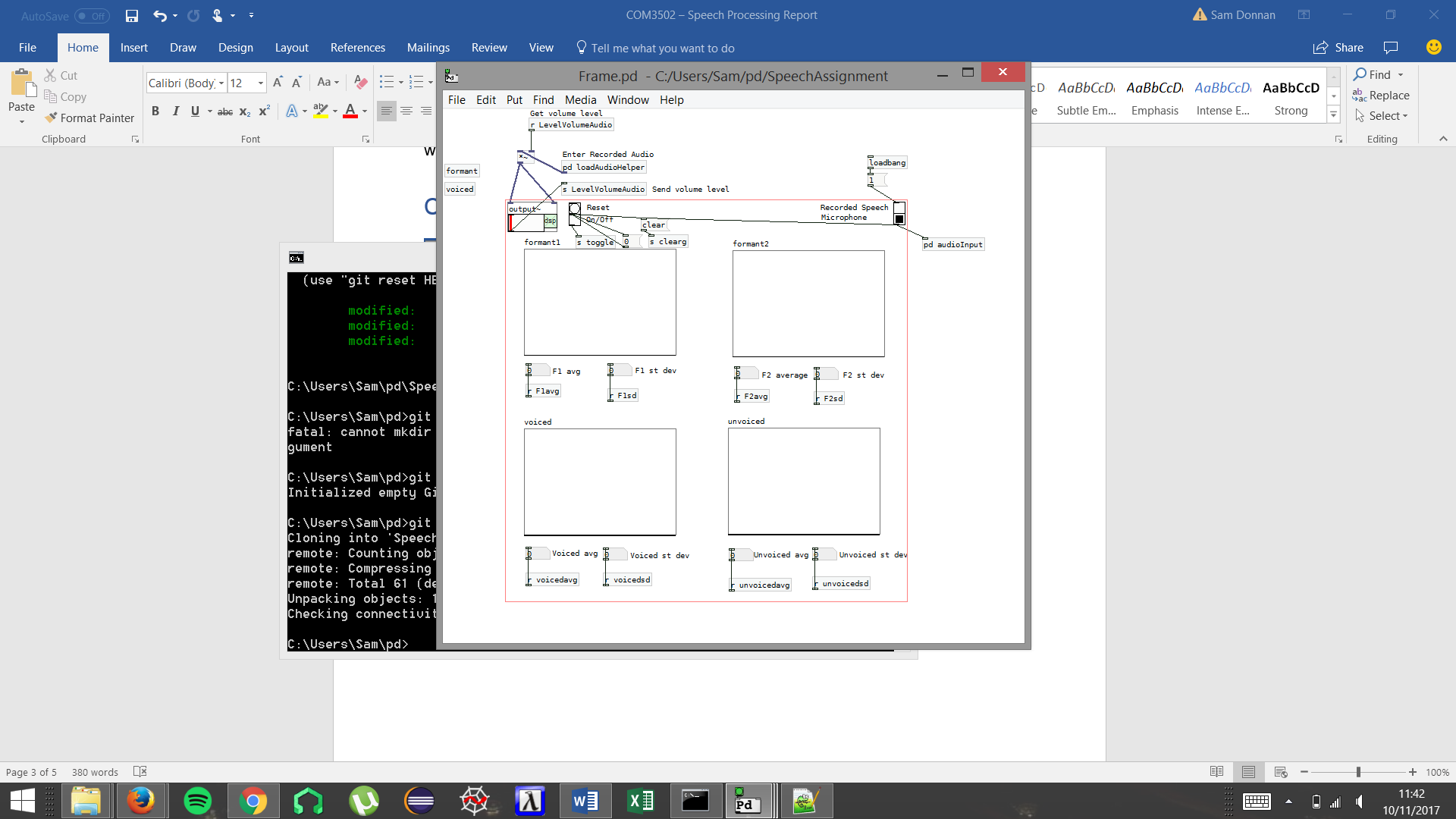
<https://en.wikipedia.org/wiki/Formant>

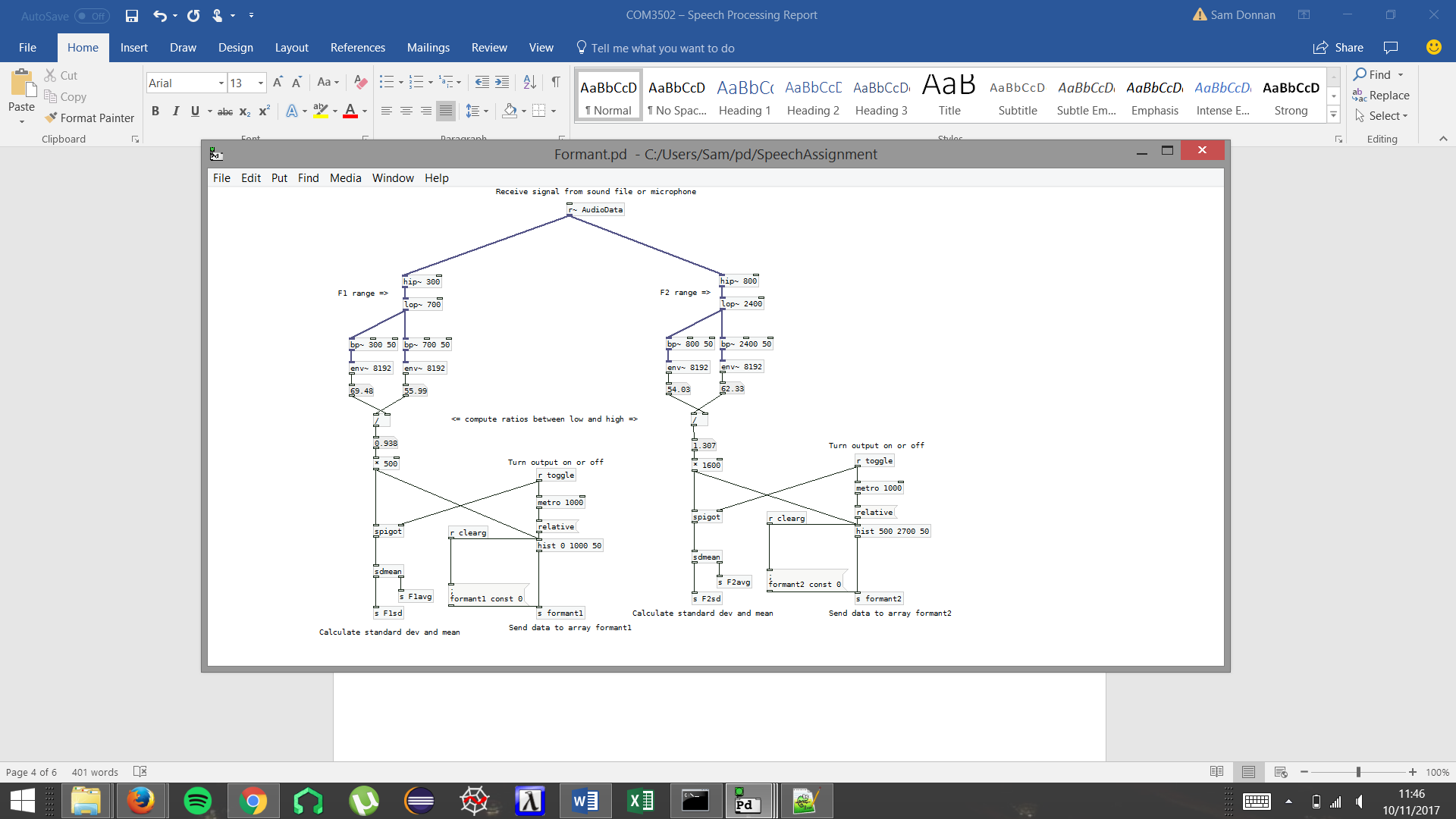
## Phonetics

To ensure reliable tests with voiced and unvoiced sounds were conducted, the media suite in the diamond was used for to its relatively soundproof qualities. In the ‘voiced.pd’ file, there is a threshold value. This value was changed and tests were conducted using live voice samples and from various sound files, specifically ‘speech.wav’. The value was changed so that when the DSP was running and there was silence the values for the voiced and unvoiced cancelled each other out. We then tested this value with voiced and unvoiced samples to ensure it was working correctly.

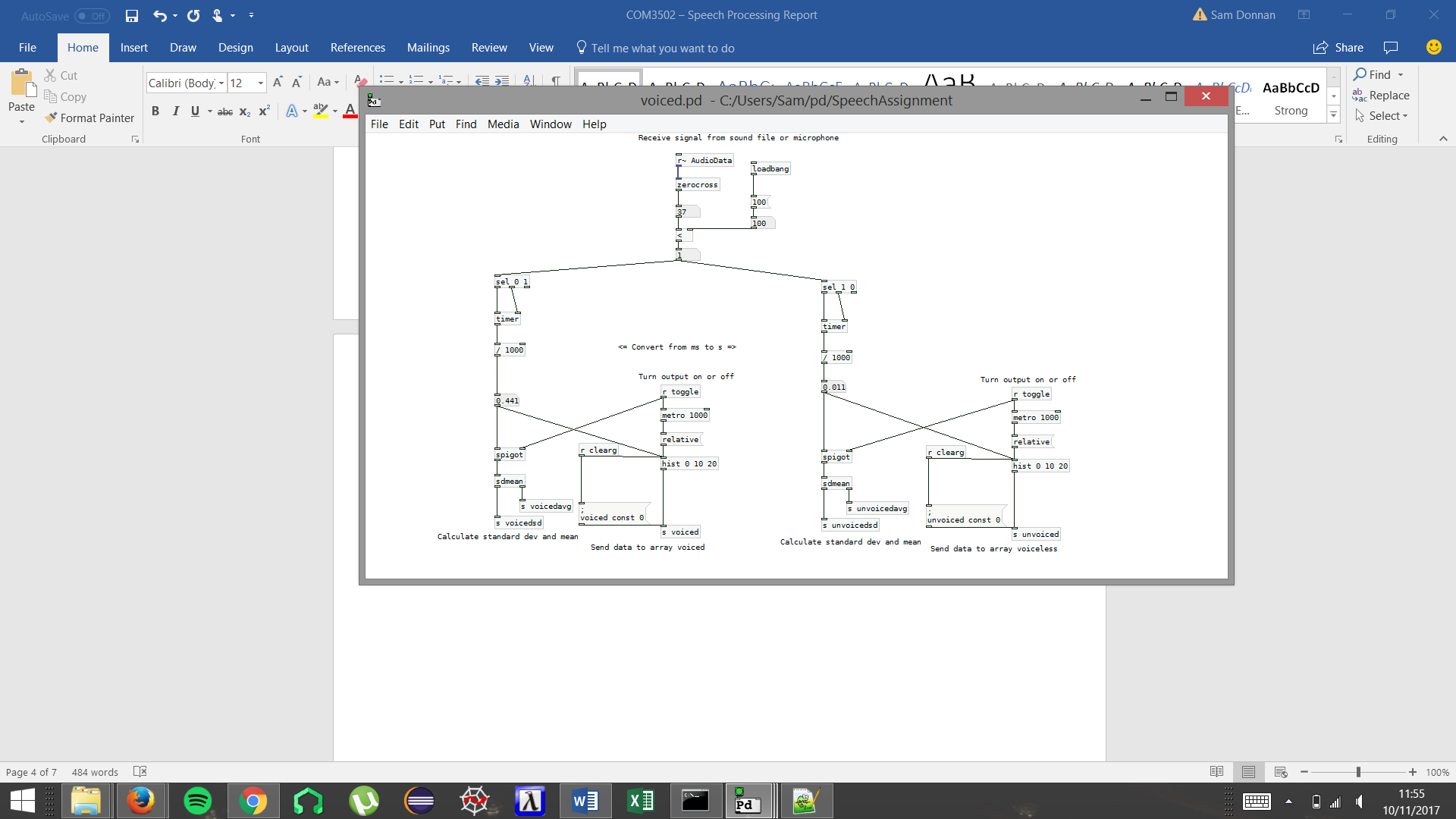
# Code Design

This is the main window used as an interface for the Assignment. It contains all the graphs for formants and phonetics, as well as the corresponding averages and standard deviations. Included at the top are toggles for the DSP, resetting the graphs and switching between input from a sound file and live audio.

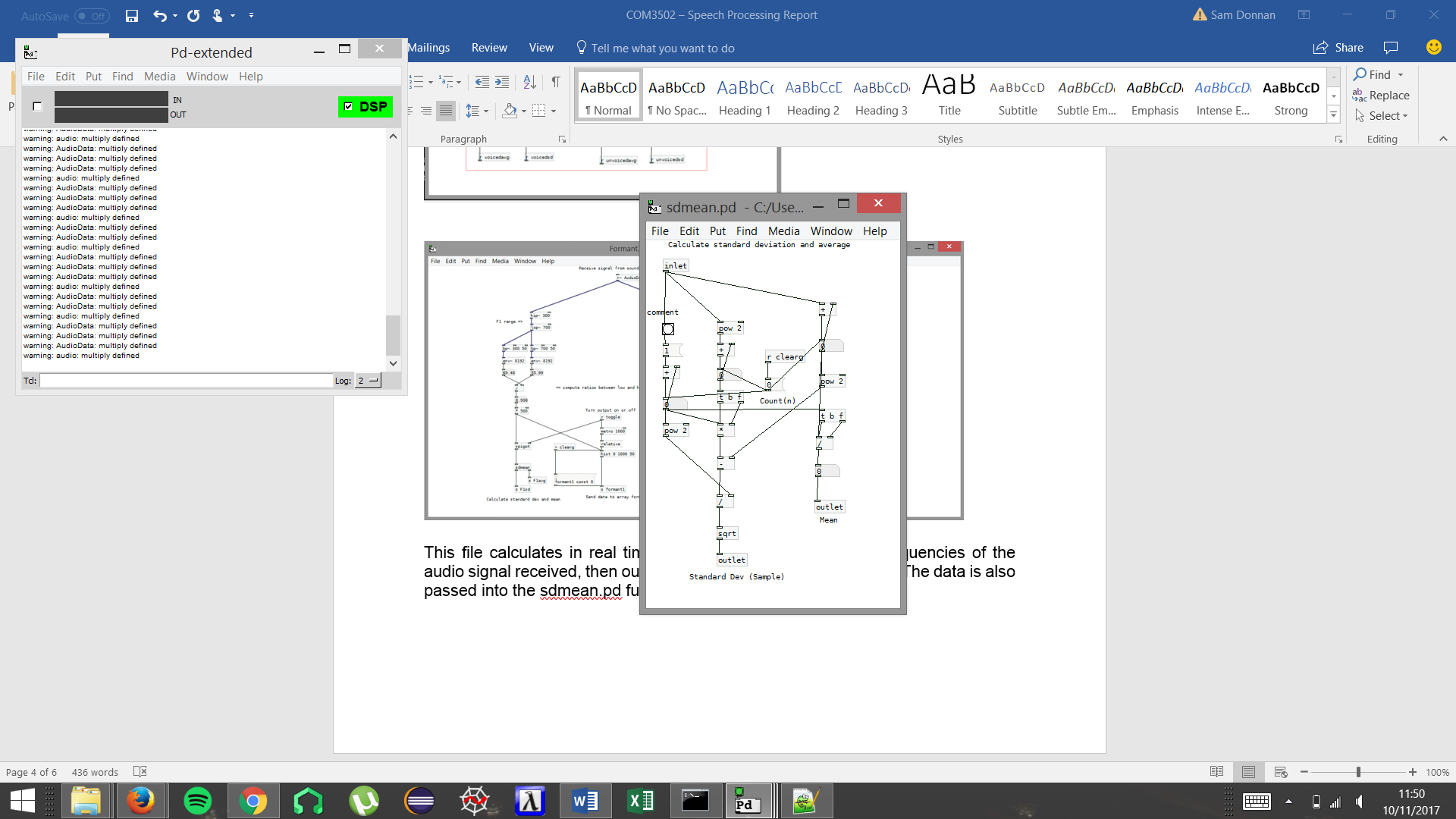
This is the frame for the main window interface that collects all the data and sends it to the relevant graphs.



This file calculates in real time the first and second formant frequencies of the audio signal received, then outputs this to the graphs in the frame. The data is also passed into the sdmean.pd function.



As with the formant file this file takes in audio data. The data is passed to the zerocross.pd function which returns a value that is then compared to a predefined threshold value to determine whether the sound is voiced or unvoiced. The duration of time the signal remains in either state is used as input for the voiced and unvoiced histograms as well as the sdmean.pd function.

This file is a function that takes numerical data and calculates the mean and standard deviation which is passed to the left and right outlets respectively. Until the function is reset it stores the previous input and uses the new data to update mean and standard deviation values.

# Results

1. How much speech (in seconds) do you need in order to obtain stable estimates of the parameters?

After conducting various experiments, it was found that it takes on average 7 seconds of speech to obtain stable estimates of the parameters.

1. What are the average durations of voiced (V) and unvoiced (UV) segments in the utterance:  “*She had your dark suit in greasy wash water all year*”?

Voiced: 0.56, 0.463, 0.578, 0.248, 0.232, 0.213, 0.234, 0.274

Unvoiced: 0.081, 0.072, 0.075 0.059, 0.064, 0.064, 0.044, 0.071

1. What are the average F1 and F2 values for the four ‘cardinal’ vowels [i], [a], [ɑ] and [u]?

[i]: f1 - 451.1, f2 - 1954

[a]: f1 – 679.2, f2 -1450

[ɑ]: f1 -

[u]: f1 – 470.1, f2 – 1710.4

<http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/filter.html>

https://itunes.apple.com/gb/app/spectrumview/id472662922?mt=8