Matlab for Voice Recognition

Raymond Fleming
Spring 2017

Use of Voice Recognition

- Voice Recognition is everywhere in the modern world.
- Personal Assistants like Siri, Alexa, and Google Home
- Onstar
- Automated Phone Systems

Uses of Fourier Transforms

- Since computers have become much more powerful at a much lower cost, new uses of Fourier transforms (specifically FFT/DTFT) have come about.
- Voice Recognition
- Antares Autotune

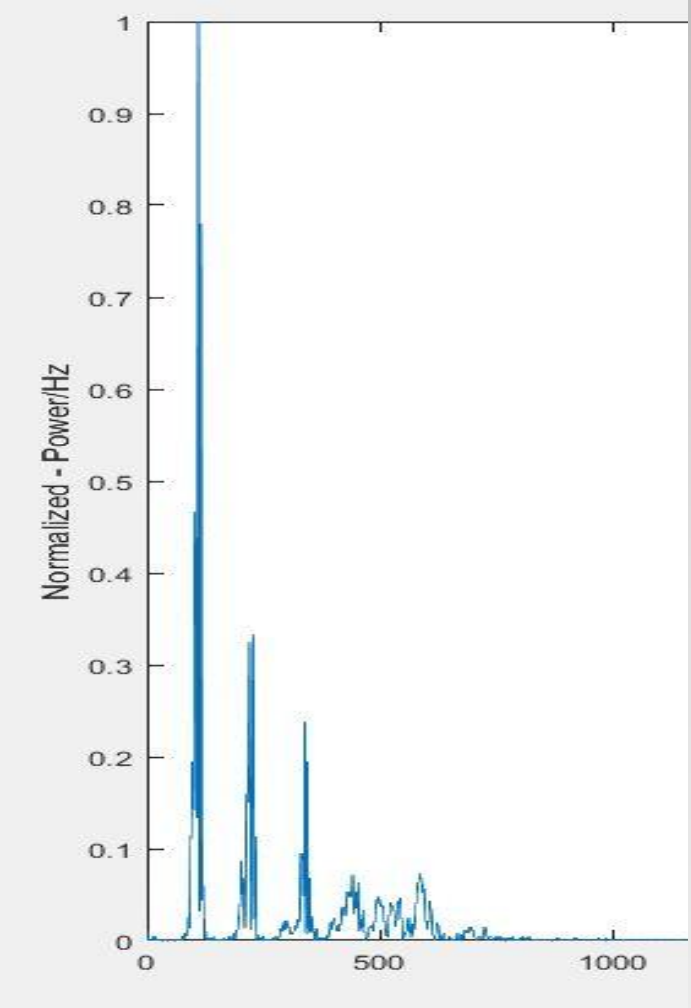
Distinguishing Yes and No

- For this project, the goal is to distinguish between a person saying “Yes” and a person saying “No”.
- This was accomplished using a power spectrum analysis based on a Fourier transform.

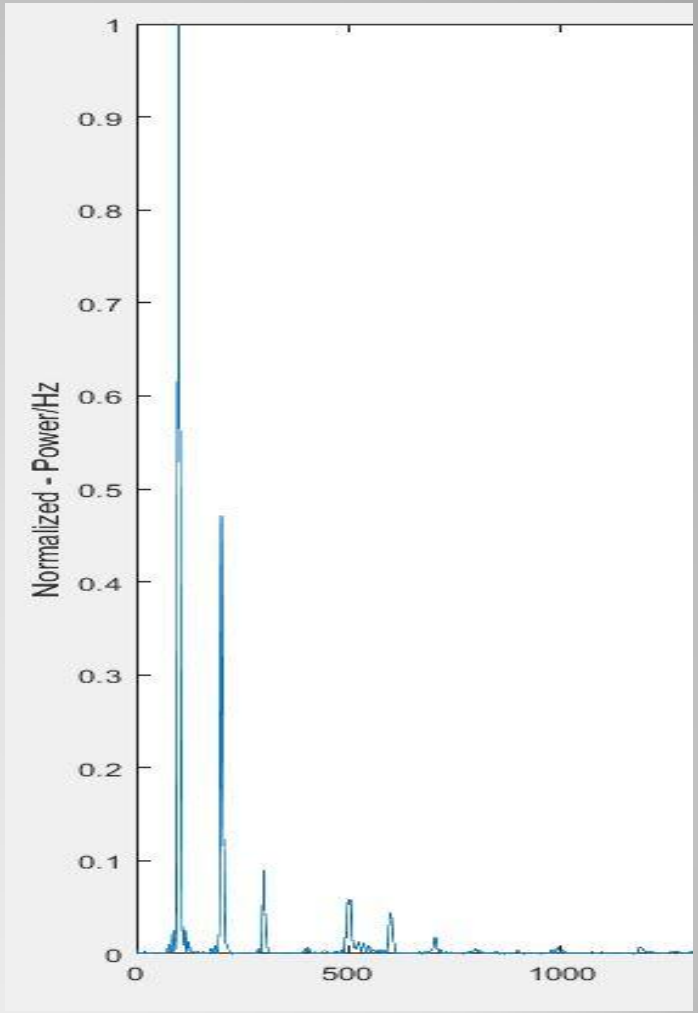
Seeing the Difference

- After viewing several plots of the power spectrum density for both “No” and “Yes” it was apparent that “Yes” has many more peaks in the upper range.
- These peaks could be used to distinguish between the words.

Yes


Normalized - Power/Hz
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
500
1000

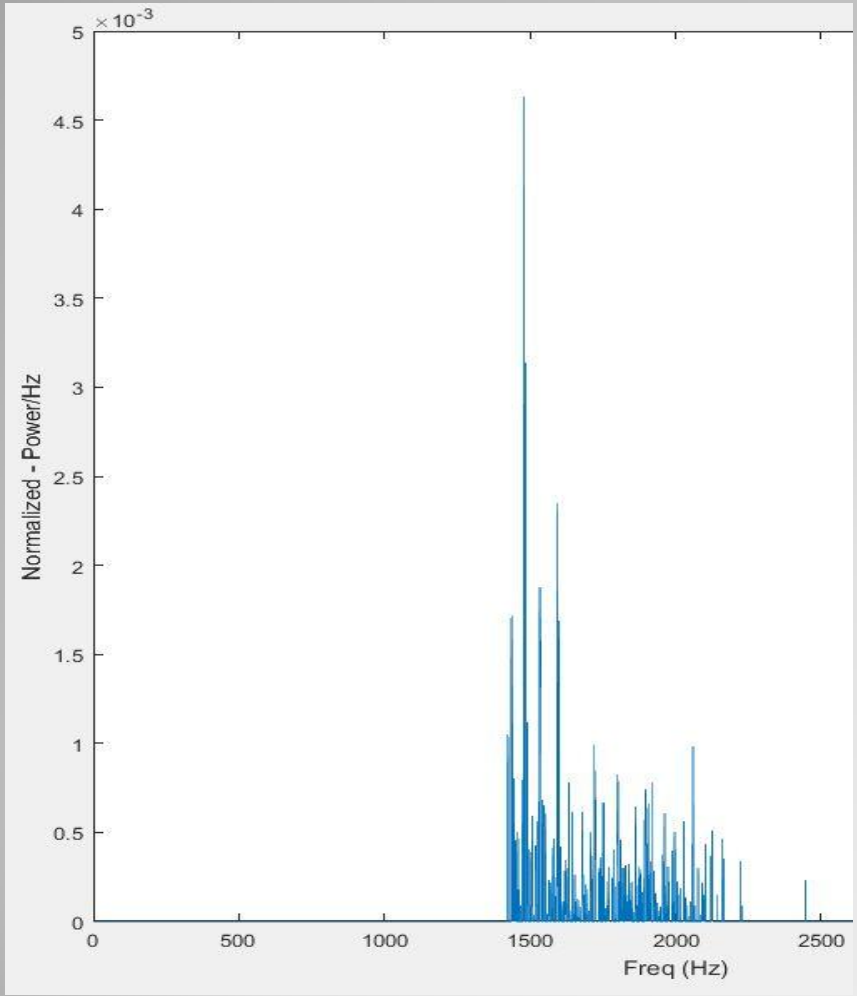
No


Normalized - Power/Hz
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
500
1000

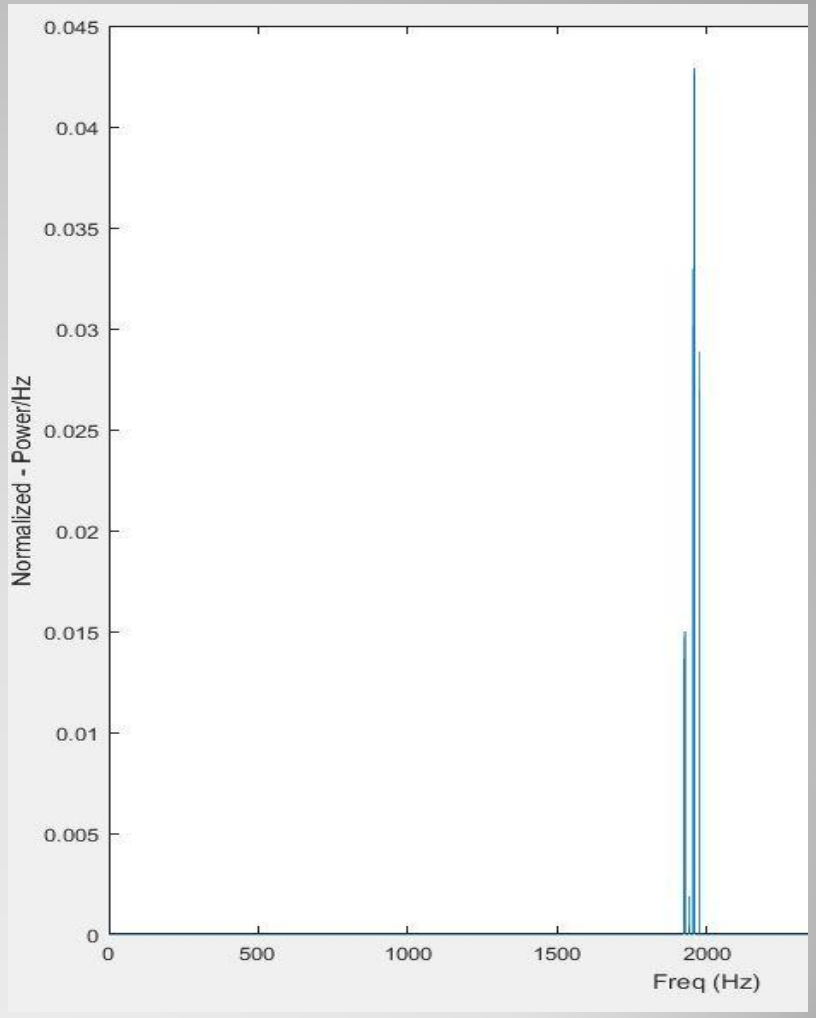
Filtering

- Not all data collected is actually needed.
  - The key data is in the upper range, so all lower frequency values can be dropped.
  - The noise is unnecessary, only actual peaks are needed. This can be filtered out as well.

Filtered Yes



Filtered No



Making the Determination

- Once the input is processed, the number of peaks is analyzed and compared to a stored value.
- If there are more peaks than the stored value, the input is “Yes”. Otherwise it is “No”.

Three Tuning Variables

- Freq_Threshold: This is the cutoff point for the frequency, all lower frequencies will be zeroed.
- Power_Threshold: This is the power level of noise. The higher it is, the more peaks get removed. It should be high enough to remove noise, but low enough to keep data.
- Peaks_Threshold: Value of the number of peaks. Higher will be “Yes” lower will be “No”.

Tuning

- The algorithm has to be tuned for an individual. To do this, view the power spectrum density values to determine where the frequency difference appears to be and set the Freq_Threshold to around this number.
- Increase or decrease the Power_Threshold until the data looks clean, but without any data loss.
- View the number of peaks for a number of yes and no trails, then pick a number above the “No” range but below the “Yes” range. There will be outliers, throw them out.

- After tuning, a 90% accuracy rate was observed over 20 trials.

Trial	Test Input	Test Output
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	Yes
6	No	No
7	No	No
8	No	Yes
9	No	No
10	No	No

Trial	Test Input	Test Output
11	Yes	Yes
12	Yes	Yes
13	Yes	Yes
14	Yes	Yes
15	Yes	Yes
16	No	No
17	No	Yes
18	No	No
19	No	No
20	No	No

Trial	Test Input	Test Output
21	Yesno	Yes
22	Test	Yes
23	Home	No

Results

Conclusion

- Using a Fourier Transform to differentiate between a small set of words is an accurate and fast way to perform speech recognition.
- Its likely that a least in part, this is how both systems like Onstar as well as some automated telephone switchboards work.