Alex Kontogiorgos-Heintz
HS361/NE360
Lab 8: Due 11-13-13

Abstract:

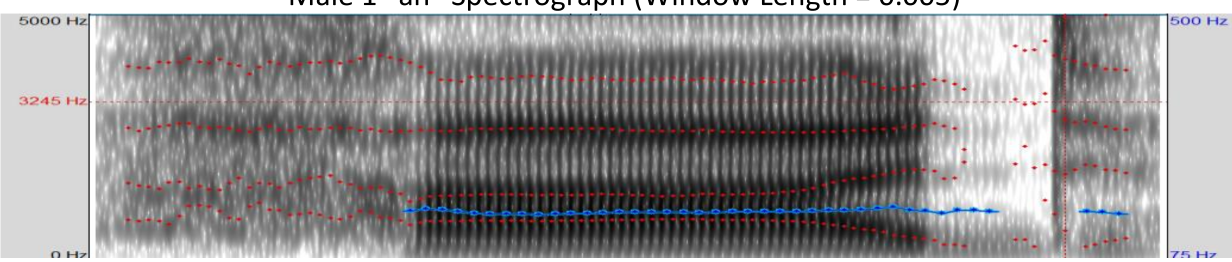
In this lab we experimented with speech analysis, specifically the application PRAAT which creates spectrographs based on .wav file inputs. PRAAT also has features that helps ascertain formant values based on the spectrographs. For inputs we used a set of prerecorded vowel sounds from an open source database. We also looked at the original Peterson and Barney vowel data in MATLAB, and compared the vowel triangles it produced to the vowel triangles produced from PRAAT.

Results:

In part 1 of the lab we used PRAAT to create spectrographs based on .wav file. We used the spectrographs to approximate the formant values for the different recorded vowel sounds. In part 1.1 we looked at the first male's sound files.

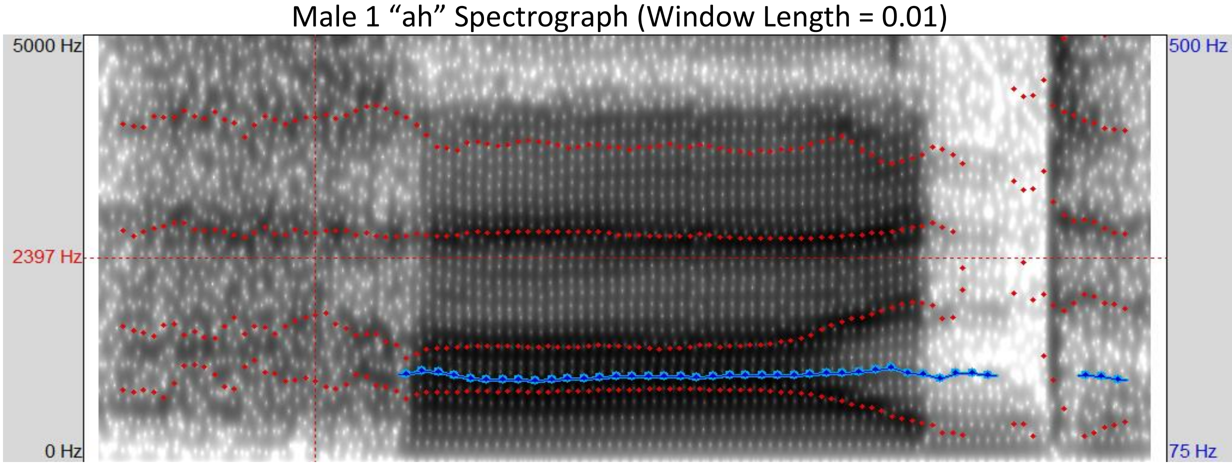
- The estimated formants lay along the peaks of the frequency spectrum; the dark horizontal lines on the spectrograph.

Male 1 "ah" Spectrograph (Window Length = 0.005)



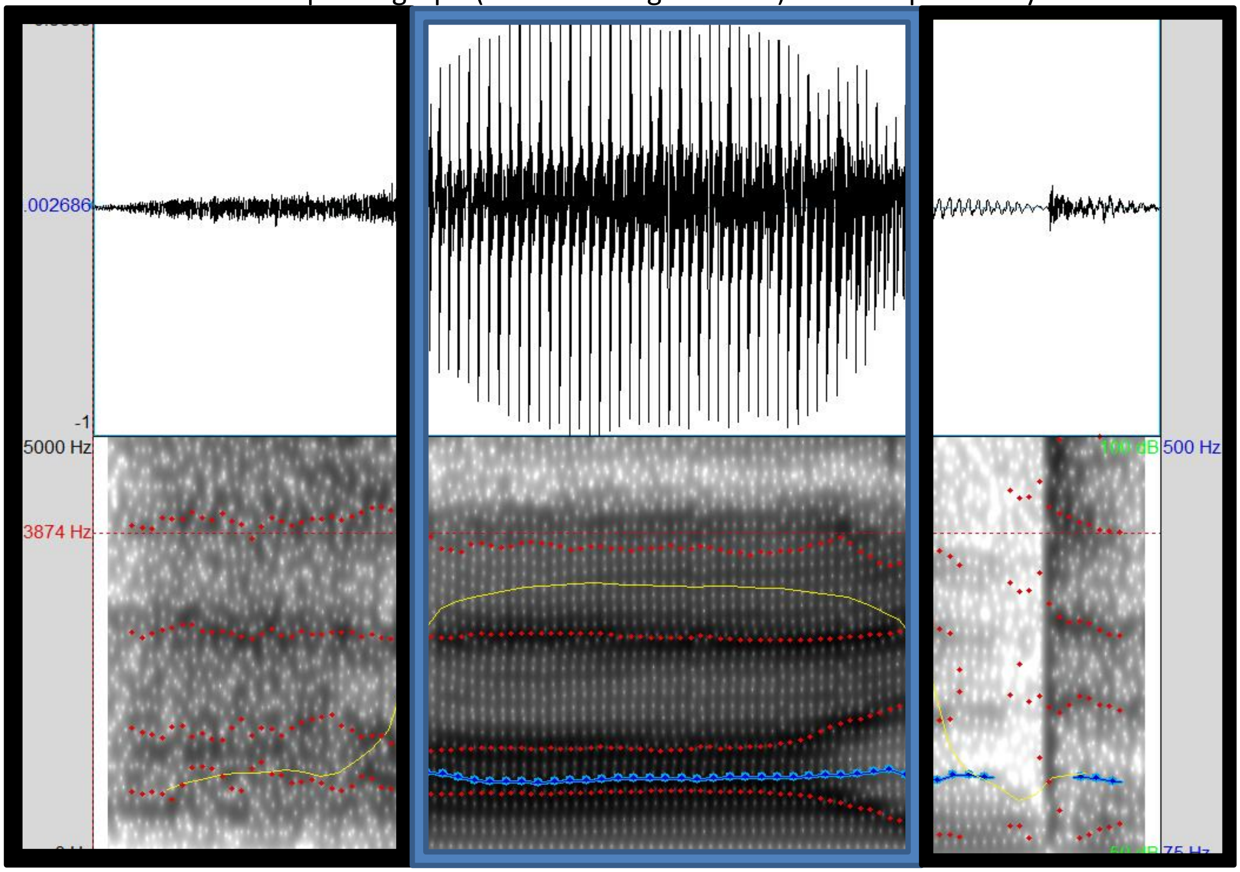
- It is a more narrowband filter, and the frequency domain has a higher resolution. This is because the Fourier Transform done on the wave form can closer approximate the correct frequency when it is given a smaller window to act on.

Male 1 "ah" Spectrograph (Window Length = 0.01)



- The time domain signal has a higher amplitude during the vowel sound, and the spectrograph has darker bands, indicating more power. The left box is the "h" sound, the middle box is the vowel sound, and the right box is the "d" sound.

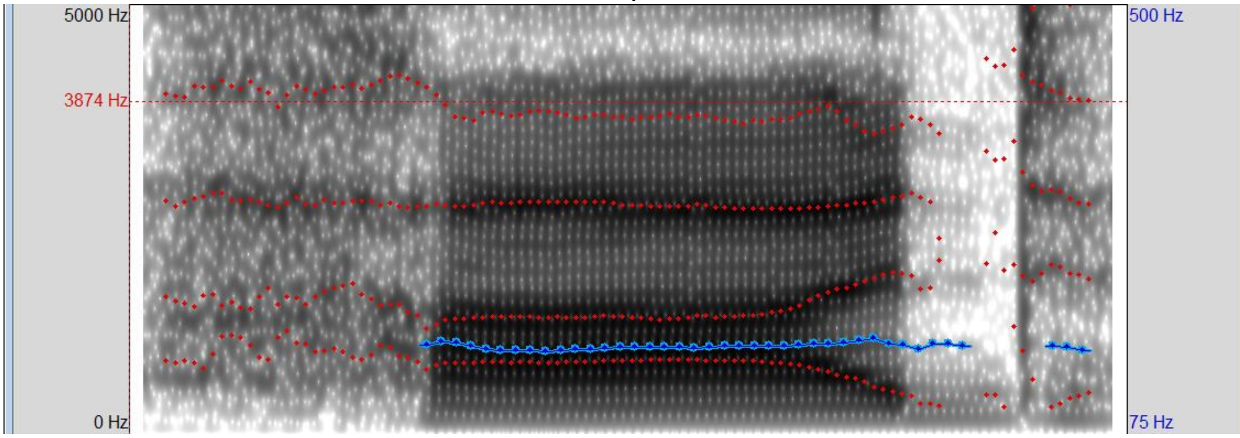
Male 1 "ah" Spectrograph (Window Length = 0.01) With Amplitude by PRAAT



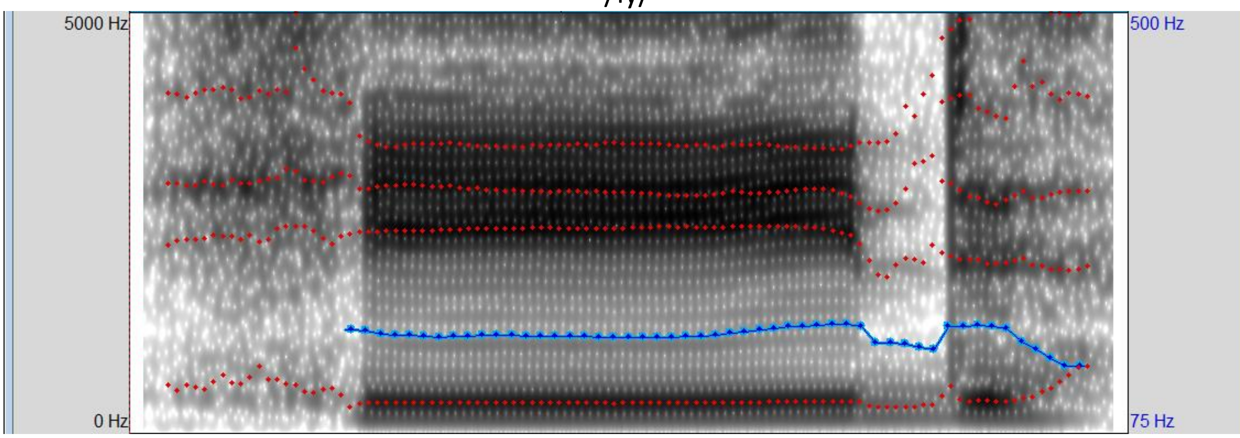
- The speaker has a voice about 20% higher than the speaker that produced the data for Figure 1.

Phonetic Symbol	**F1 (Hz)**	**F2 (Hz)**	**F3 (Hz)**
/ah/	847	1349	2654
/iy/	348	2423	2866
/uw/	375	928	2681

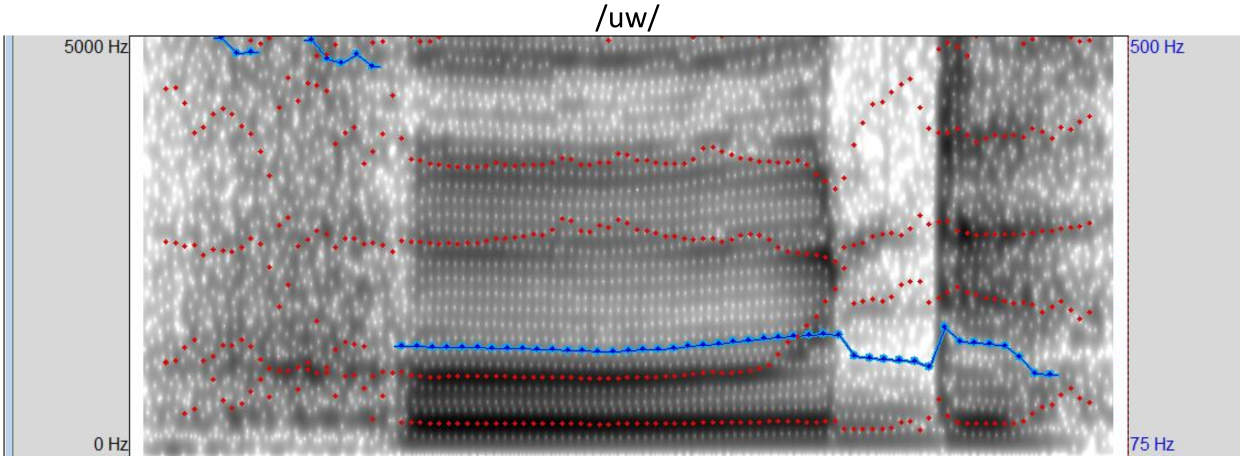
/ay/



/iy/



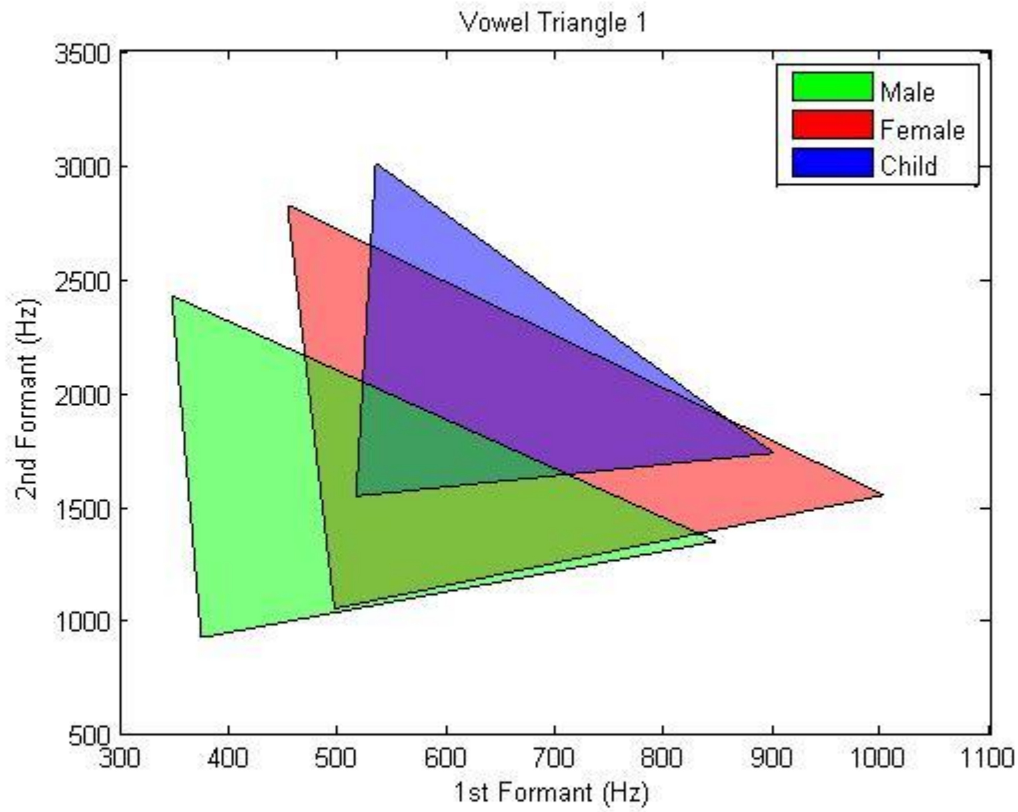
/uw/



In part 1.2 of the lab we compared the data from the previous section (analysis of a man's voice) to the voice of a woman and a child.

- The formant triangle of the man is generally lower pitch than the woman and child. The woman has a lower pitch voice than the child, and the child generally has the highest pitch voice. This is caused by the size of the larynx in the different individuals. Men typically have larger larynxes than women, and young children would have small ones that have not yet fully developed.

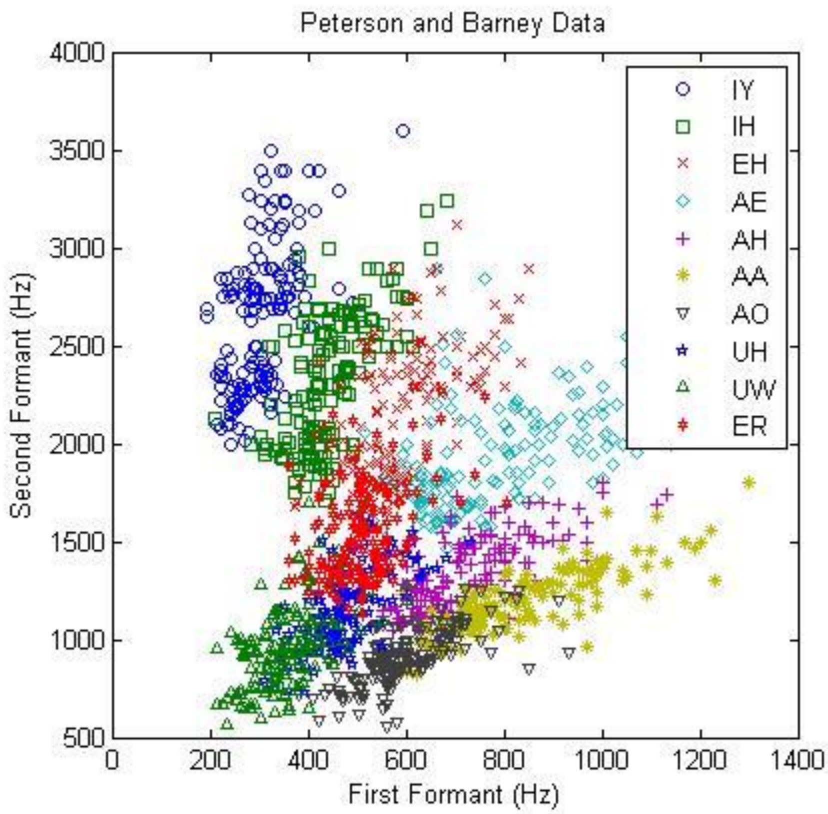
Vowel	F1 (Hz)	F2 (Hz)
Man	--	--
/ah/	847	1349
/iy/	348	2423
/uw/	375	928
Woman	--	--
/ah/	1001	1551
/iy/	454	2820
/uw/	497	1055
Child (Boy)	--	--
/ah/	900	1735
/iy/	535	3003
/uw/	517	1547



In part 2 of the lab we looked at the Peterson & Barney database of vowel formant values. The purpose of this is to show the strong clustering of vowels by their first two formants.

- The reason it is important to detect features that are predictive of the speech sound is because a given sound could represent multiple different words/syllables based on the speaker's voice. The fundamental frequency of the sound could be an additional feature since it could help to differentiate between speakers, since, for example, men tend to have lower pitch voices than women. Neurons in the auditory stream that are sensitive to fundamental frequencies might therefore be incorporated into this process.

- The shape of the clusters seem to be a linear projection from a central point just right of the origin. The F1, and F2 values from the previous part of the lab seem to lay on the further out area of the clusters, but still defiantly within.



- The results are very consistent, the percent correct was always just shy of 70%, the easiest to guess were always /iy/ and /ao/, and the hardest were always /eh/ and /ah/. These results (for hardest and easiest) are not unexpected since /iy/ and /ao/ are clustered in areas by themselves, where as /eh/ and /ah/ are in areas with lots of overlap. Since there are 10 vowel classes the percent correct if it were pure chance would be .5^10. Given this very high rate of correct predictions (relative to pure chance), the first two formants seem incredibly important in analyzing speech.

Conclusion:

In this lab we worked created formant triangles for different voiced speech sounds, and related them back to a commonly used data set. To create these triangle we made spectrographs based on a time series of a sound wave, and extrapolated the formant values from them. The results of the lab showed the extensive importance of the first two formants for speech analysis.