Harmonic Content of Vowels

Phys 498 POM

John Nichols

I. Introduction

Timbre is defined by Wikipedia as the quality of a musical note or sound or tone that distinguishes different types of sound production, such as voices or musical instruments. Timbre is one of the aspects of sound and music that most intrigues me. How can I tell the sound of a violin, trumpet, saxophone, or voice apart? How can my ears distinguish a speaker on the other end of a telephone?

After a bit of research online and in books, I discovered it is a combination of sound *envelope* as well as the *harmonic content* of the source. Envelope refers to the attack and decay of the sound. Harmonic content, which is what will be discussed from here on out, is the relative volumes of the overtone frequencies. In the human voice, consonants seem to deal with the transient changes in envelope while vowels are mostly dependent on harmonic content.

The human voice is also fundamentally different from other instruments in its construction. One model created to explain the way the voice works is the Source-Filter Model. This says that the vocal cords (source) produce a periodic sound that is changed by the position and resonances of the vocal tract (filter). The fundamental difference of the voice is that the vocal chords can produce any note between its continuum of notes and these notes do not necessarily correspond to the resonances of the vocal tract. In a wind instrument, you can produce the fundamental resonance of the length of the instrument and by changing the length, you can modify what those notes are. The voice is not like this. Humans can modify the resonances of the filter and the fundamental frequency independently. The Source-Filter Model is detailed more completely on the University of New South Wales acoustics website.

This ability to vary resonances independently of pitch is what gives voice the variety of vowels we have. These resonances are called formants. The first few formants are mostly independent of one another and controlled by mouth position. The first formant (F1) is largely determined by mouth positioning (open/closed), while the second formant (F2) is determined by tongue position (front/back). Researchers at the University of New South Wales claim that the first two formants are the most important parts of vowel identity.

II. Recording and Software

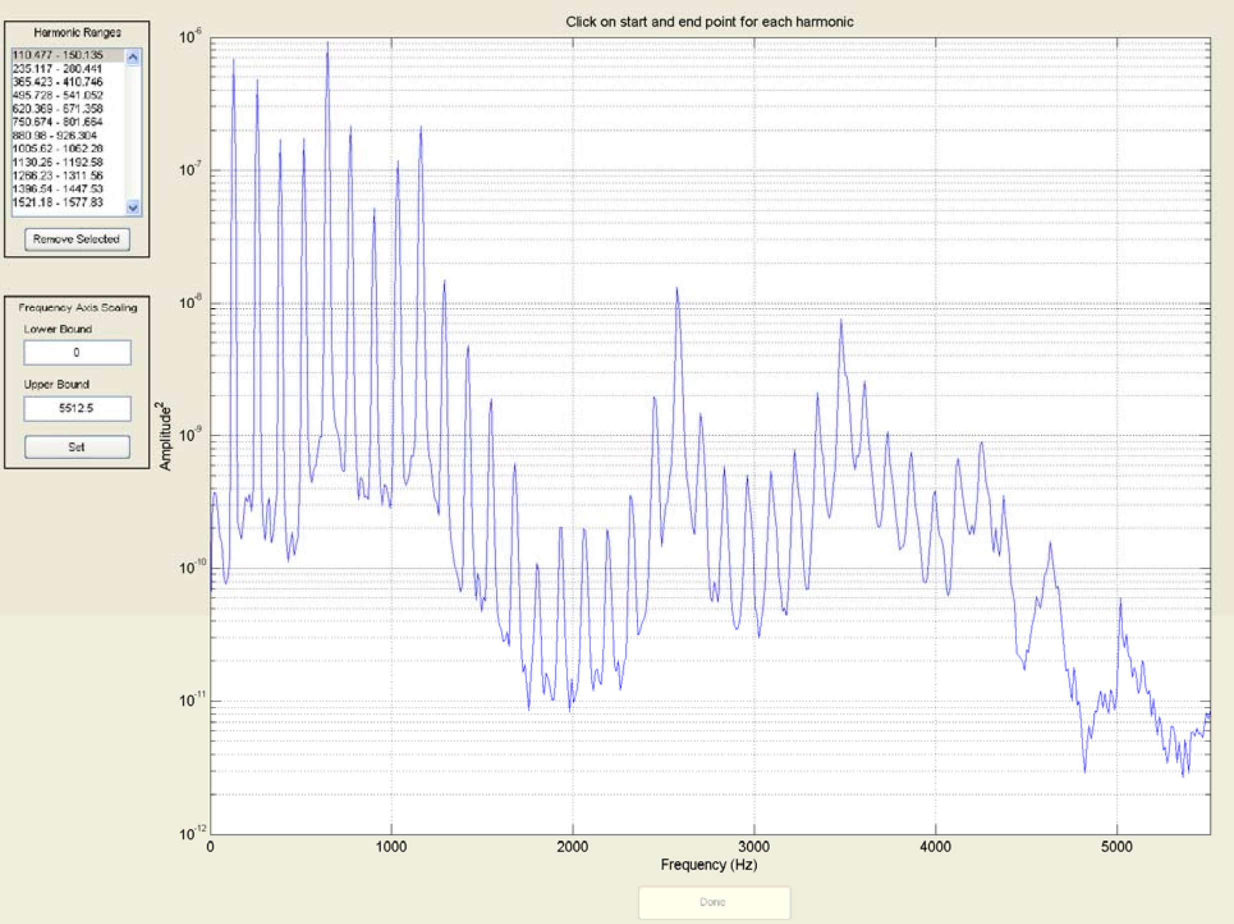
I recorded a number of samples from two subjects. The majority of the samples were recordings of my own singing, but a few (primarily the alternate techniques) were from a good friend of mine, Zach. The primary sounds recorded were five vowels, [a] (ah), [e] (eh), [i] (ee), [o] (oh), and [u] (oo). I recorded these on four different notes, C3, G3, C4, and G4. Zach was recorded on all five vowels, only on C3 and C4. All samples were recorded with a 24-bit digital sound recorder using a dynamic microphone.

The sound files were transferred to a computer to be analyzed using the wav_analysis.m program (also known as the Wave Analysis Toolbox) designed by Joe Yasi. The Wave Analysis Toolbox was my primary tool in all my analysis on this project. The Wave Analysis Toolbox (WAT) provides a lot of information about the input wave file. It gives a frequency-domain plot of amplitudes, as well as amplitude, frequency, and phase information of each harmonic you ask it to. Unfortunately, because the human voice fluctuates in pitch a relatively large amount (which causes phase to change drastically), the phase information was inconsistent, so I largely ignored it. The thing that interested me most was the amplitude of the harmonics and how these related to the sound.

III. Data Analysis

As stated in the previous section, the information I looked at mostly was the amplitudes of the harmonics. I initially thought that vowel information would ultimately be contained in the first few harmonics of the sound in a combination of amplitude and phase. But as I said before, the phase information was inconclusive. And surprisingly, in the first five or six harmonics, the vowels were not entirely discernible and their harmonic content changed drastically when the fundamental pitch changed. When my data confused me, I looked to other research and discovered the information about formants I describe above.

Armed with a new outlook, I expanded the number of harmonics analyzed by the program. Now the patterns that before seemed entirely random were given some amount of predictability. I learned that the vowel [i] is not characterized by low third through ninth harmonics with loud tenth through fourteenth harmonics but by a general sound level increase between 2 and 3 kHz. Here were the patterns for the similarities behind vowels. We tune our mouths to bring out certain frequencies above others, and the absolute location of these formants are what create distinct vowels. Here are the harmonic spectra for the [a] vowel:


Click on start and end point for each harmonic

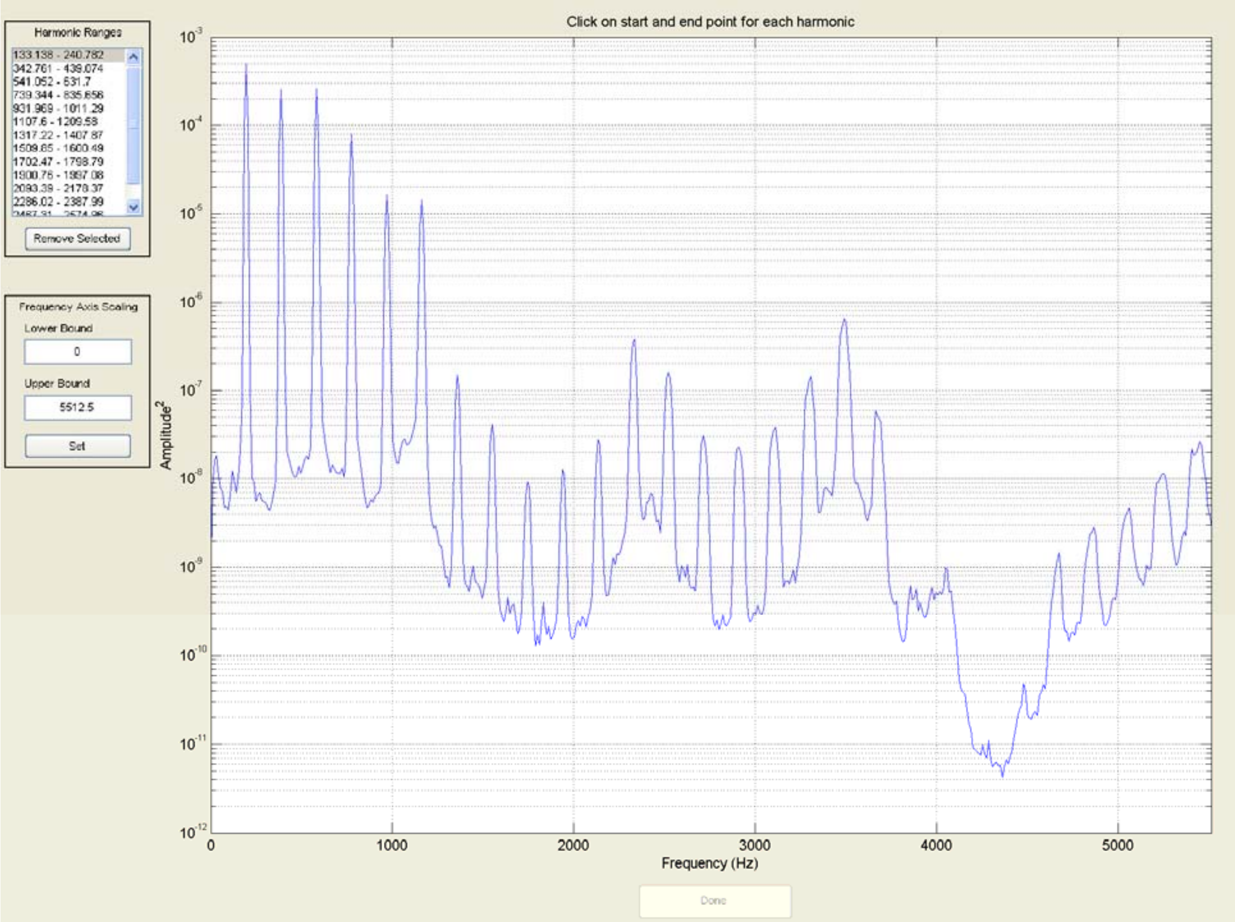
Harmonic Ranges

110.477 - 150.135
235.117 - 280.441
365.423 - 410.746
495.728 - 541.052
620.369 - 671.358
750.674 - 801.664
880.98 - 926.304
1005.62 - 1062.28
1130.26 - 1192.58
1266.23 - 1311.56
1396.54 - 1447.53
1521.18 - 1577.83

Frequency Axis Scaling

Lower Bound: 0

Upper Bound: 5512.5


Click on start and end point for each harmonic

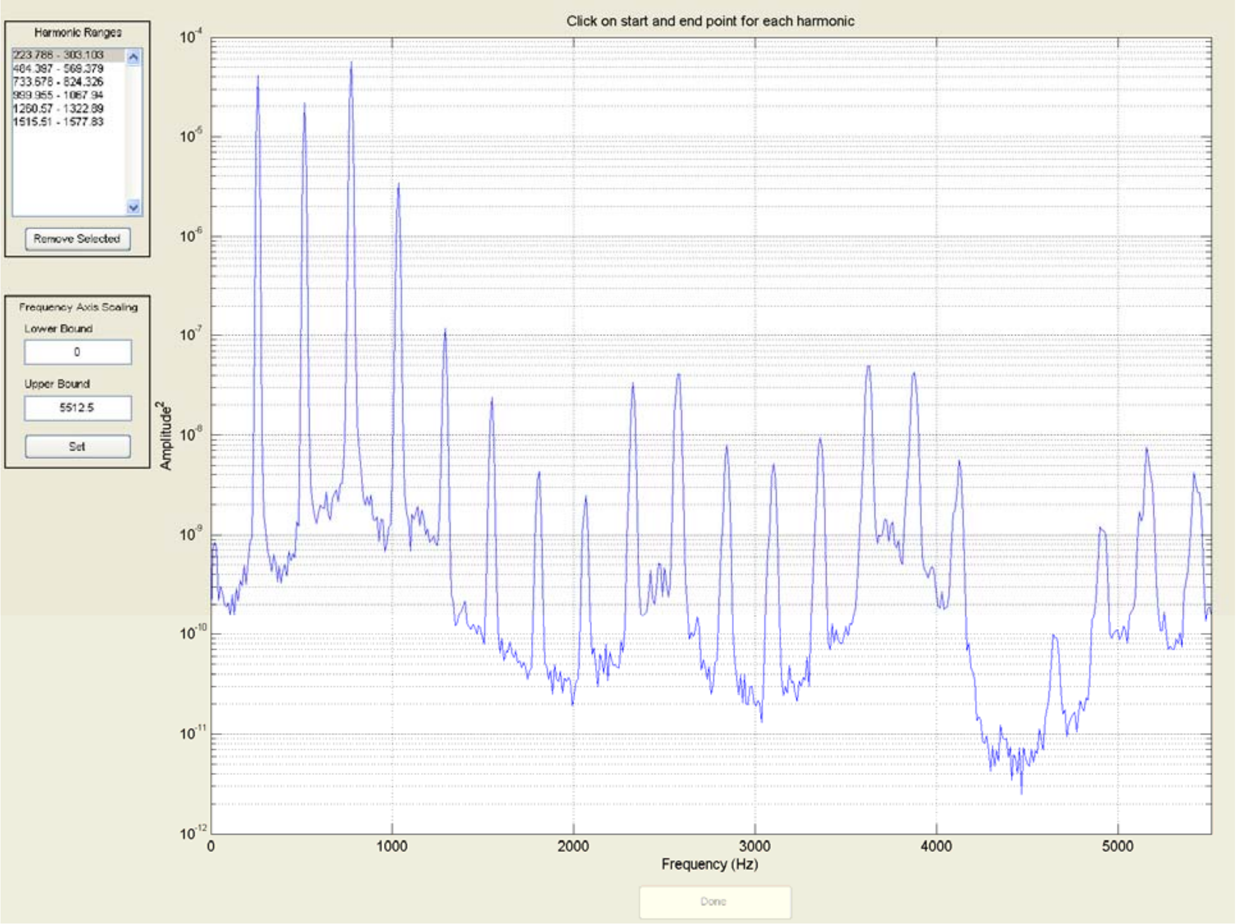
Harmonic Ranges

133.138 - 240.782
342.761 - 439.074
541.052 - 631.7
739.344 - 835.656
931.969 - 1011.29
1107.6 - 1209.58
1317.22 - 1407.87
1509.85 - 1600.49
1702.47 - 1798.79
1900.76 - 1997.08
2093.39 - 2178.37
2286.02 - 2387.99

Frequency Axis Scaling

Lower Bound: 0

Upper Bound: 5512.5

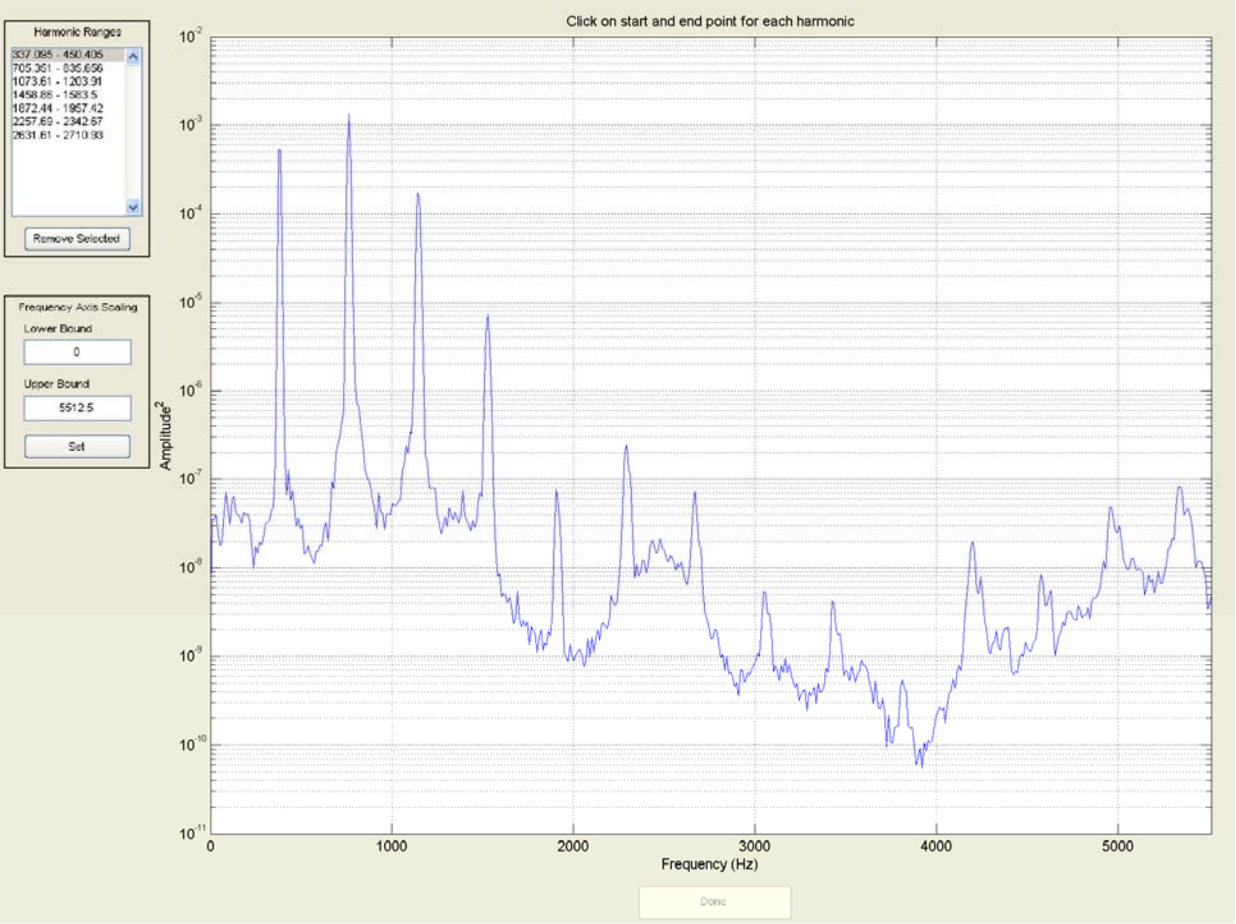

Harmonic Ranges

223.786 - 303.103
484.397 - 569.379
733.678 - 824.326
999.955 - 1067.94
1260.57 - 1322.89
1515.51 - 1577.83

Frequency Axis Scaling

Lower Bound: 0

Upper Bound: 5512.5


Harmonic Ranges

337.095 - 450.406
705.351 - 835.656
1073.61 - 1203.91
1458.86 - 1583.5
1872.44 - 1957.42
2257.69 - 2342.67
2631.61 - 2710.93

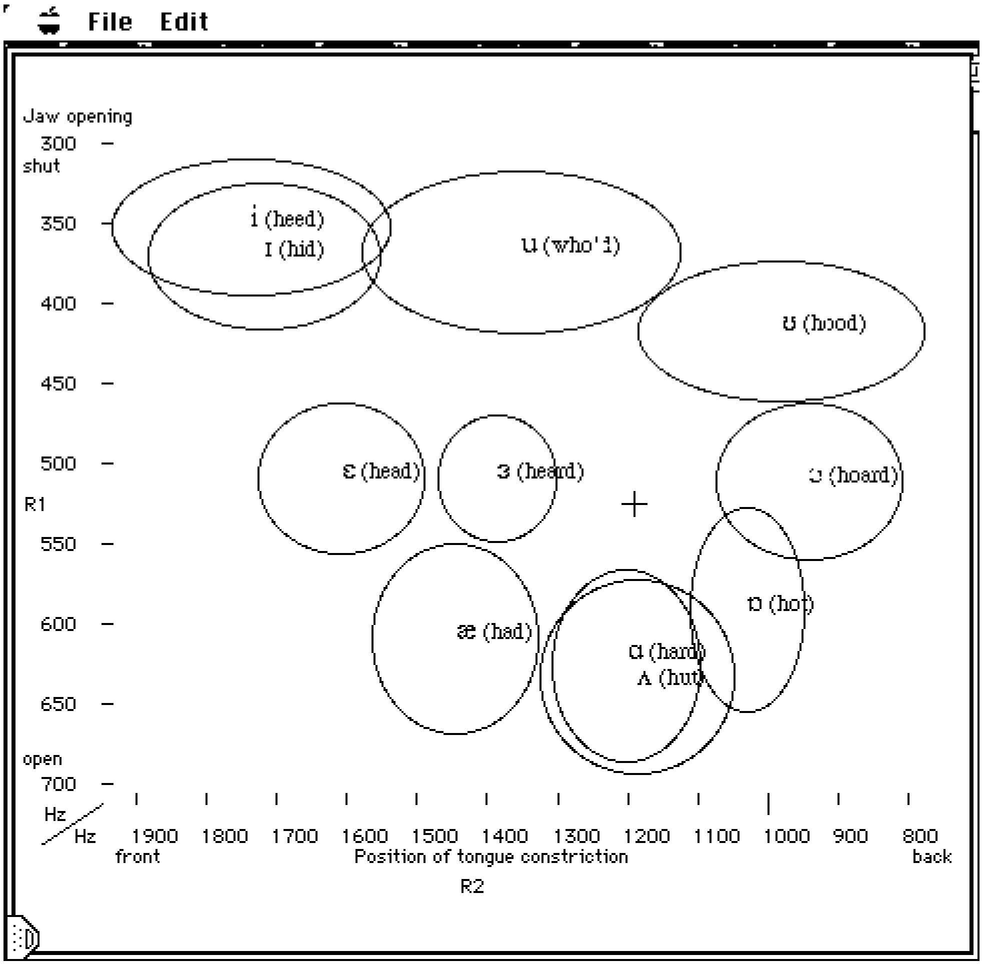
Frequency Axis Scaling

Lower Bound: 0

Upper Bound: 5512.5

These are the four samples for [a]. Notice the big bumps in the graphs around 750 Hz, 1200 Hz, 2500 Hz and 3500 Hz. These are the first four formants and they uniquely identify an [a] vowel. In the image below (borrowed from the UNSW music acoustics website), you can see the general locations of the first two formants for many vowel sounds.

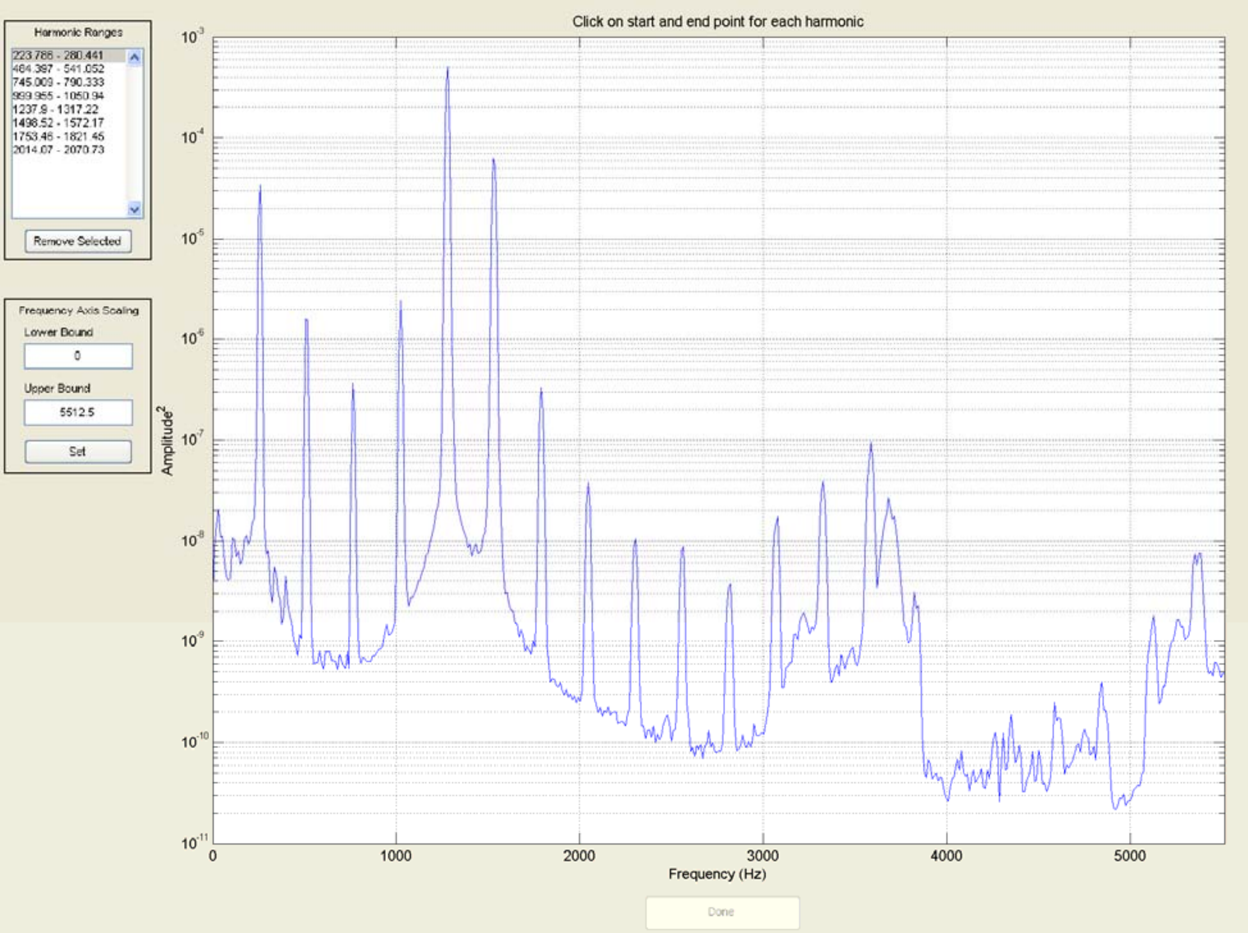
(Keep in mind, these are Australian pronunciations below.)



I also analyzed whispered speech. The lower formants present in the whispers matched well with the formants of the sung vowels. The higher spectrum varied more.

IV. Further Topics of Interest

Another thing I looked into are alternate methods of singing. One such method is overtone singing, a method in which the singer intentionally tunes the vocal tract to specifically bring out a higher harmonic, sometimes making it louder than the fundamental pitch. Overtone singing can sound like a person is singing one note while whistling a higher note at the same time. Another interesting style of singing is low throat singing. Some people suggest that in this method, the singer puts enough air through their vocal chords to vibrate the false vocal chords an octave lower than the note they are singing. This gives a very full but almost dirty low rumbling sound. This very low fundamental is very weak, but it does give extra harmonics all the way up the spectrum. Two pictures will be shown below, first the overtone singing followed by the low throat singing.

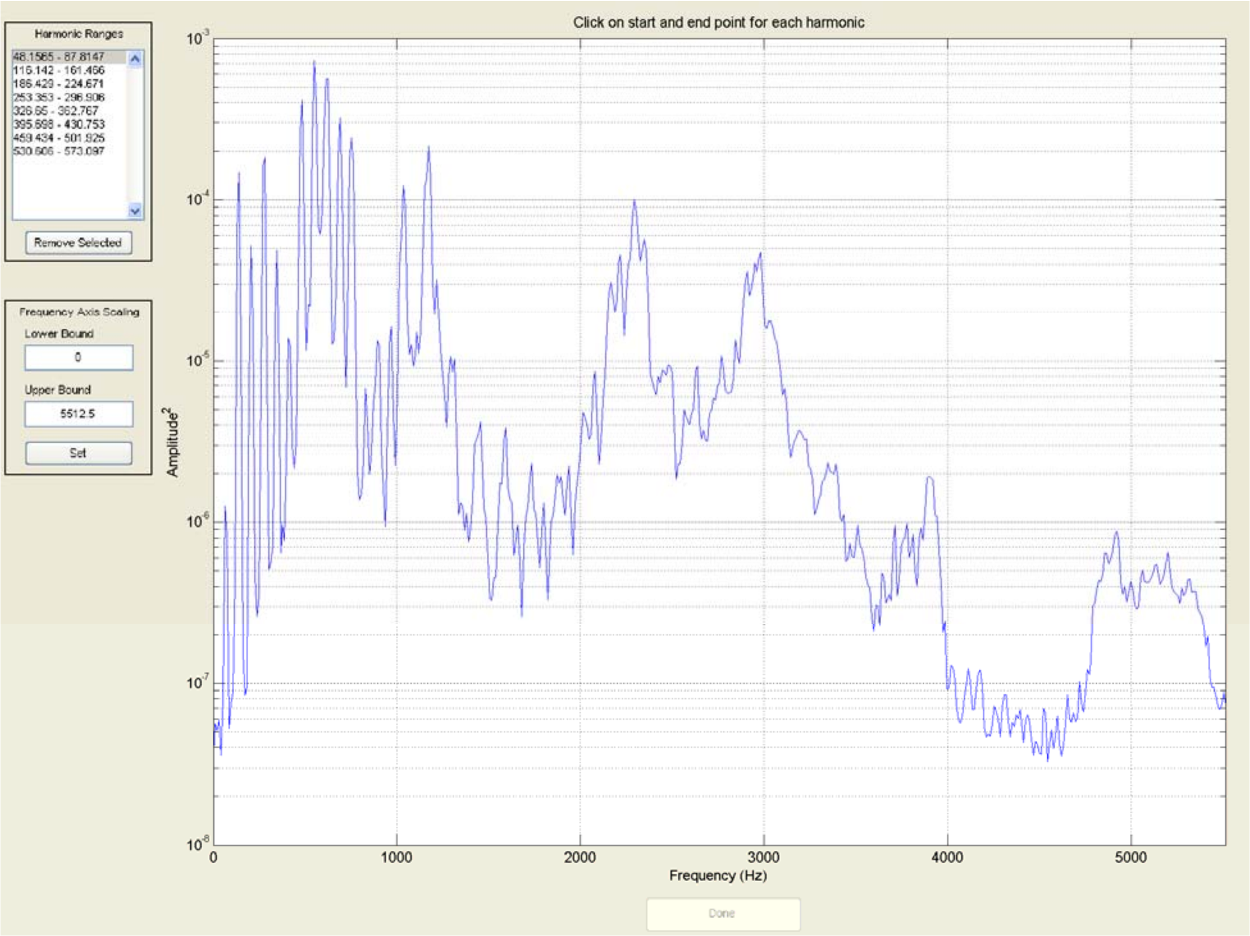


```
223.786 - 280.441
484.397 - 541.052
745.009 - 790.333
999.955 - 1050.94
1237.9 - 1317.22
1498.52 - 1572.17
1753.46 - 1821.45
2014.07 - 2070.73
```

Lower Bound: 0

Upper Bound: 5512.5

The fifth and sixth harmonics here are very clearly amplified.



Harmonic Ranges

48.1565 - 87.8147
116.142 - 161.456
186.429 - 224.671
253.353 - 296.906
326.65 - 362.767
395.698 - 430.753
459.434 - 501.925
530.606 - 573.097

Frequency Axis Scaling

Lower Bound: 0
Upper Bound: 5512.5

The sample above is on an [a] vowel, and the formants are in similar places to our other sung examples, just with many more peaks. Also, note the very quiet, low first harmonic.

With the WAT available for my use, I did a bit of tinkering, just to see what would come of it. The two most interesting things I did with that were using a low-pass and high-pass filter on the sound. The low-pass filter obfuscated the vowels almost entirely. I filtered out the harmonics, adding one in at a time, and with each harmonic added, the vowel was more identifiable. Once the first and second formants were reached, the vowels were distinguishable, but prior to that, they were pretty similar.

The high-pass filter did not have the same effect. In fact, the high-pass filter only succeeded in making the sound file sound less bassy and more tinny. Even when the fundamental pitch was not present, the fundamental could still be heard.

The very last thing I did with my project, as a culmination of what I learned over the course of the semester, was to try to produce synthesized vowels using Sony Sound Forge 8. I added together sine waves at the various amplitudes from my analyzed vowels to create the best vowels I could. They turned out to be distinguishable from each other, but not necessarily uniquely identifiable on their own.

V. Conclusion

Overall, this was a rewarding and enlightening look into how we produce the vowel sounds we do every day. I learned many things like what primarily distinguishes vowels, but my imperfect synthesized vowels suggest there is a little more to the story than just the harmonic amplitudes (perhaps phase or frequencies that fall between the exact harmonics). I learned more about how the voice works in general, and this knowledge has changed the way I think about singing. My testing pool was not wide enough to truly draw any conclusions from my data, but my results agree with the published results of the research I have been able to find on vowels, both spoken and sung.