Special Problems Final Report

ECE 4901 - Dr. Anderson

Connor Hawley

12/18/2017

A Quantitative Approach to the Effects of Vocal Training

I began this semester's work under Dr. Anderson where our work as a group left off in the spring. To preface this report, this project began as a more tractable spinoff from Dr. Anderson's original ambition to predict a chicken coop's general health using only audio. Initially, the group began looking for data with which to evaluate healthiness, and we searched for high quality audio from the Biggest Loser and other weight loss shows to no avail. Eventually, an opportunity to record the voices of a local opera came about, so we shifted the work's focus to understanding exactly what separates practiced and untrained voices. By May 2017, we had accumulated some qualitative findings that we presented in a final deliverable to the opera house itself.

However, I felt that there was much quantitative analysis yet to be done, so I set out to do just that in August, again under Dr. Anderson's supervision. Using covarep, an open-source, Matlab based signal processing library, I began by analyzing the glottal closure instance (hereafter referred to as GCI) derivatives with the script "**analysisGlottal.m**". This first program served as infrastructure on which to build later analysis. The procedure is simple: it calculates the GCI derivative of the current signal using one of covarep's built-in functions, covarepAnalyzeDGF, normalizes the GCI derivative, then stores that result as a row in a master array. A similar procedure is applied to the raw waveforms in the "**analysisRaw.m**" script to obtain normalized results in a master array for further analysis in other scripts.

After establishing some infrastructure to facilitate more complex analysis, I worked to calculate the kurtosis of each GCI derivative waveform to look for a pattern among the practiced and untrained singers in the "**analysisGlottalKurtosis.m**" script. The procedure utilized MATLAB's built-in kurtosis function, which uses the formula $\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$. This analysis proved to be somewhat fruitful: the GCI derivative kurtosis of the untrained singers tended to be higher than the GCI derivative kurtosis of practiced singers, implying that the practiced singers' GCI derivatives are much more consistent than untrained singers, which could be interpreted as more "control" over the voice. This result is displayed in figure 1.

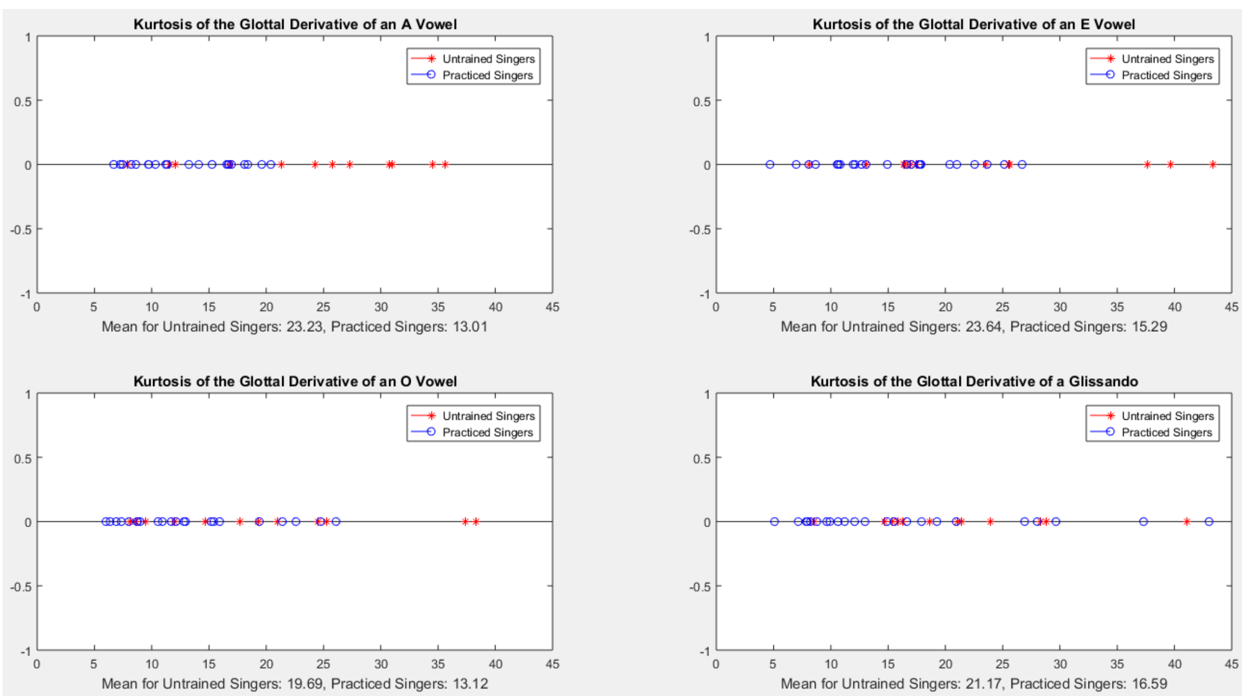

Figure 1. The GCI derivative kurtosis of practiced and untrained singers singing different vowels and a glissando.

Later, this analysis of the kurtosis of the GCI derivatives piqued my interest for the same analysis to be used on the raw audio. The script "**analysisRawKurtosis.m**" performs this task, with less suggestive results. The grouping of the untrained and practiced groups is not nearly as pronounced for the raw audio as it is for the GCI derivatives. I am still satisfied that I did this analysis, as it underscores the richness of information contained in the GCI derivative. The results can be seen in figure 2.

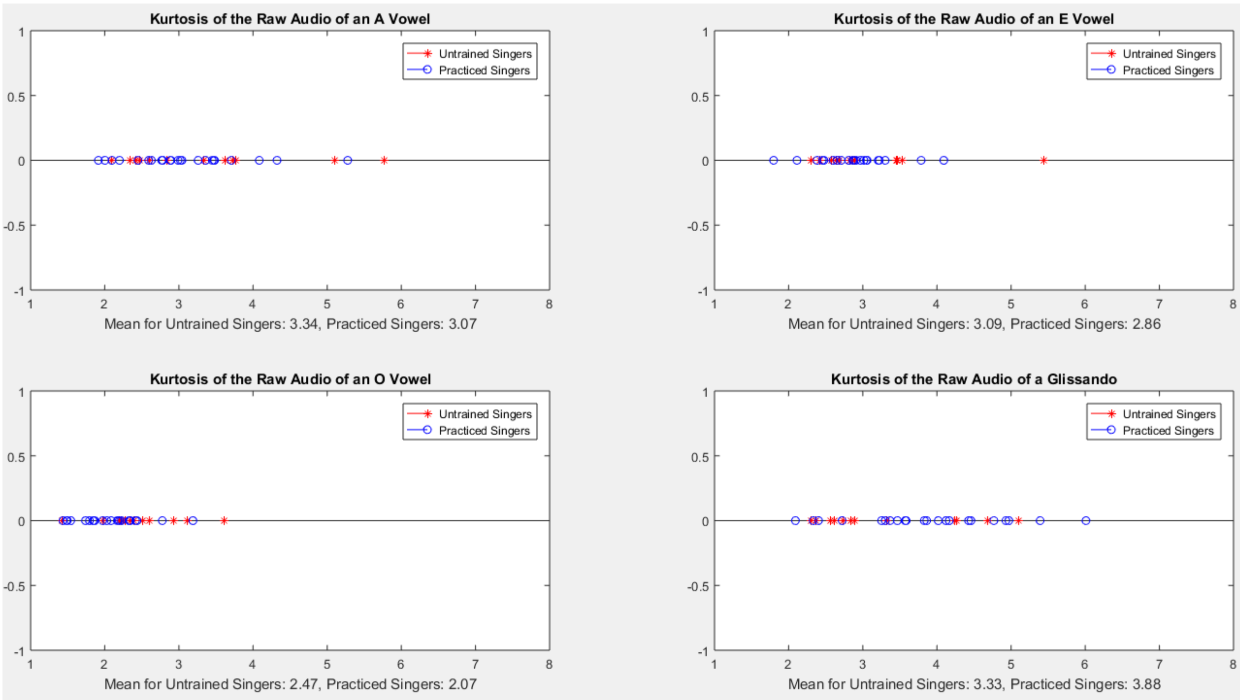

Figure 2. The raw audio kurtosis of practiced and untrained singers singing different vowels and a glissando.

The next quantitative method applied was to observe the pseudo phase space plot of the GCI derivative in the "**analysisGlottalPPS.m**" script. This procedure, inspired by concepts in chaos theory, involves creating 3-tuple coordinates of 3 samples: some sample, a sample some delay behind the first, and a sample twice some delay behind the first, $(s[n], s[n-\tau], s[n-2\tau])$. This delay is calculated with the expression $\tau = \frac{fs}{f_o}$, so each waveform's pseudo phase plot has each successive coordinate being delayed by roughly one period. Thus, all three coordinate components should theoretically be very similar.

These 3-tuples are then plotted as a curve in three-dimensional space. I expected to see that some pattern would be closely repeated to form a tight curve for practice singers, while untrained singers' patterns would appear more chaotic and random. While both the practiced and untrained singers' pseudo phase space plots appear chaotic and random, there is an important difference. The practiced singers' curves tended to cluster much closer to the origin than the untrained singers'. To verify my conjecture, I found the average distance from the origin for each singer's pseudo phase plot, plotted that for each vowel, then found the standard deviation of those averages, and the latter 2 results are recorded in figure 5. As for the pseudo-phase space plots themselves, the first 10 practiced and untrained singers' plots are documented in figures 3 and 4.

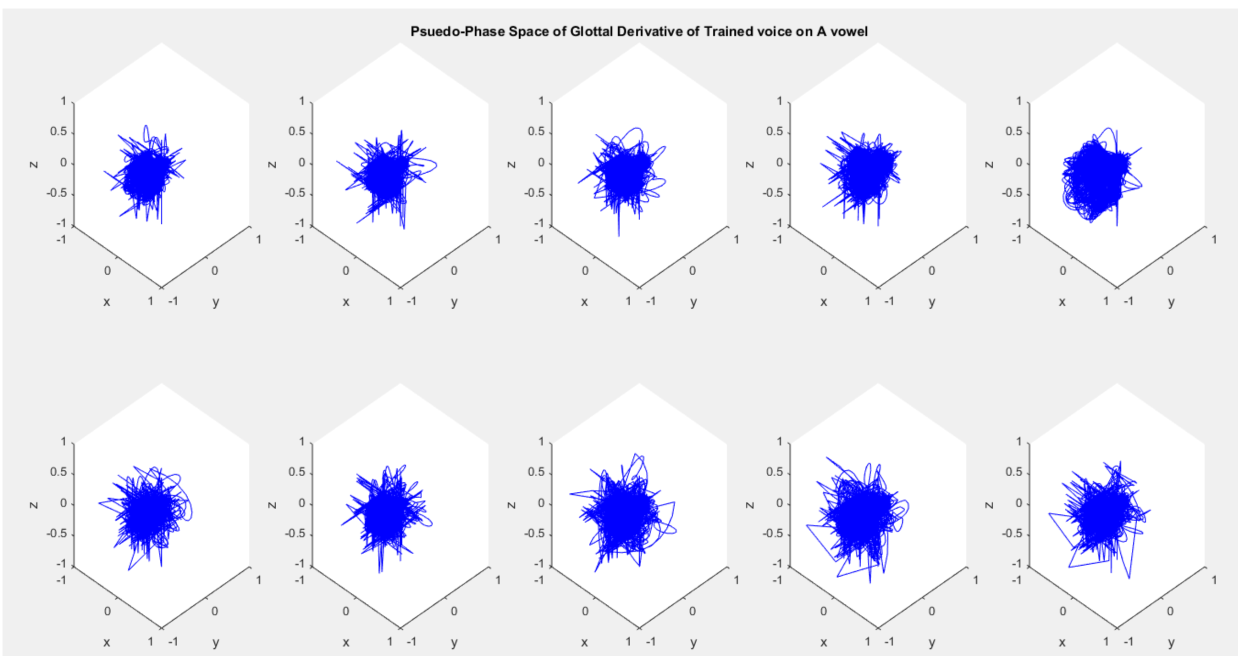

Figure 3. GCI derivative pseudo-phase space plots of practiced singers singing an A vowel.

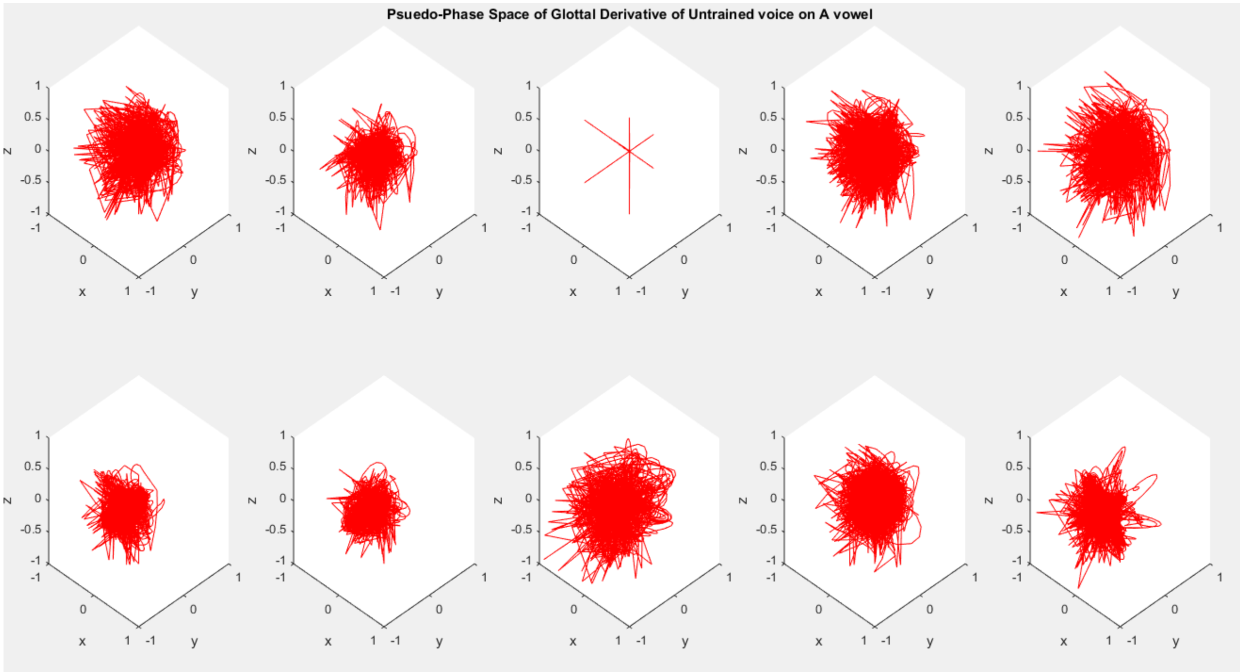

Figure 4. GCI derivative pseudo-phase space plots of untrained singers singing an A vowel.

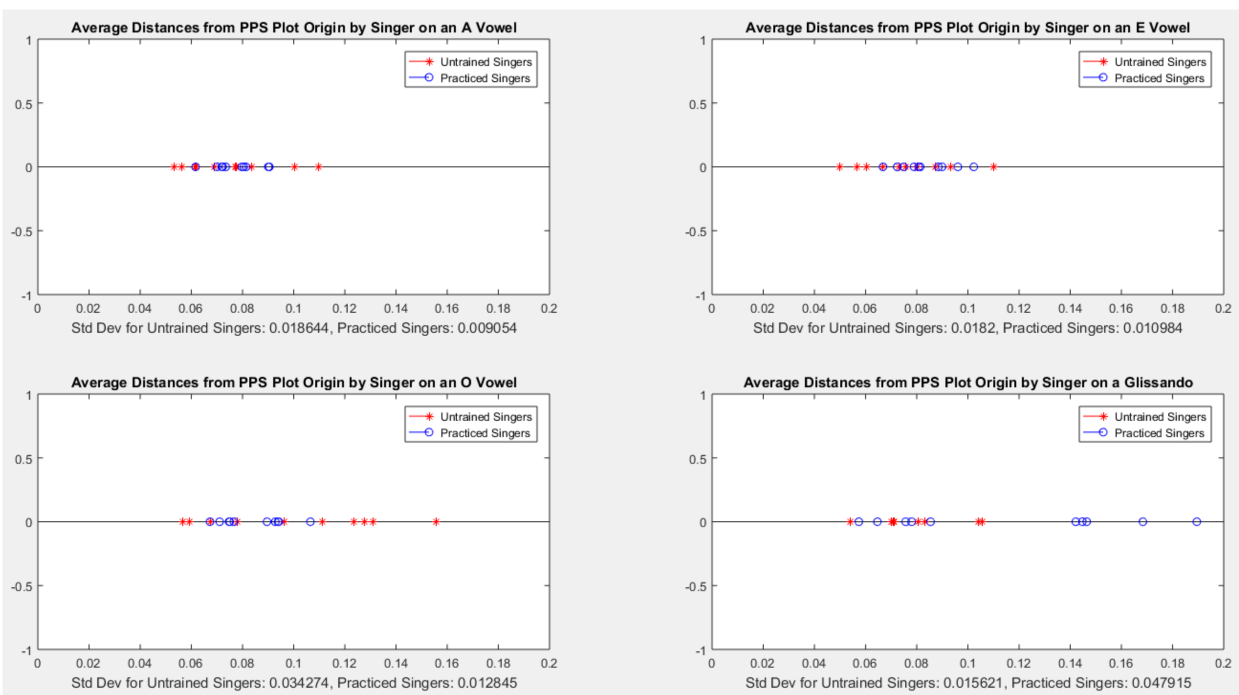

Figure 5. GCI derivative pseudo-phase space plot distance from the origin by vowel and by singer.

Pitch in the raw audio waveform was also considered. I used covarep's sum of residual harmonics algorithm to estimate fundamental frequency in the raw audio waveforms. These results are averaged over time for each vowel of each singer to gain an overall best-guess on the waveform's fundamental frequency. These averages are then plotted by vowel to give insight on the differences between practiced and untrained voices. Remarkably, the practiced singers are very tightly clustered at specific pitches, whereas the untrained singers are all over the place. This result is displayed in figure 6.

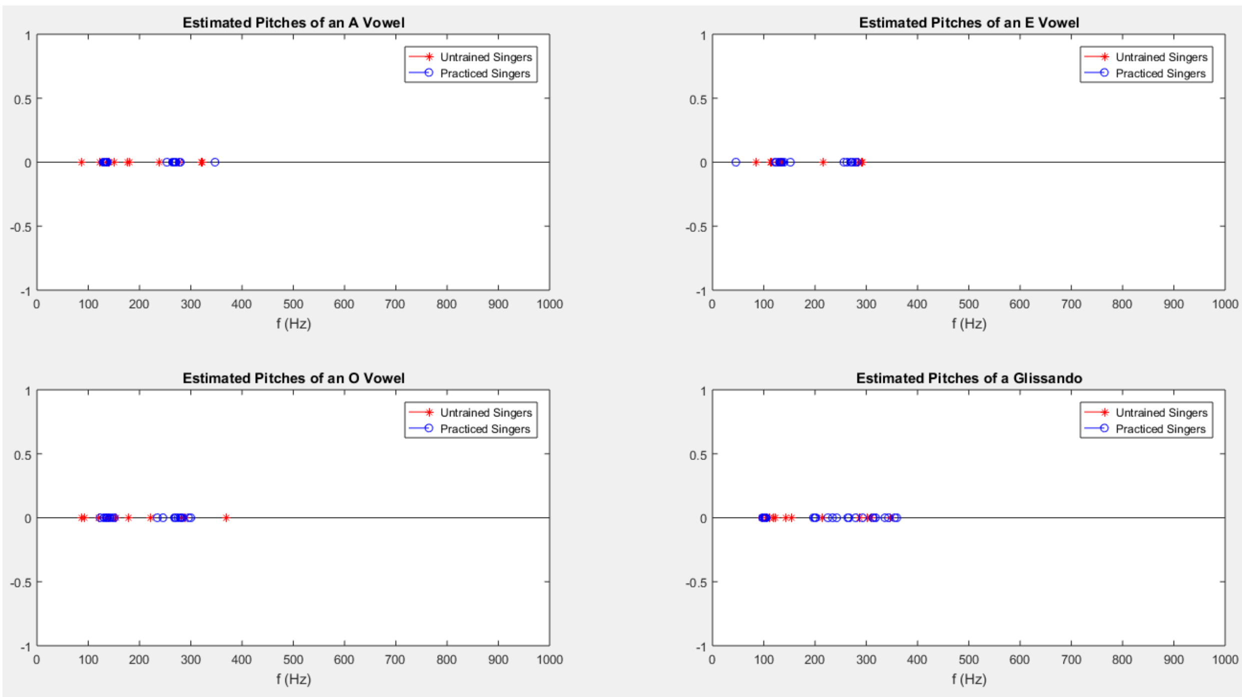

Figure 6. Raw audio SRH harmonic plots for each vowel and the glissando.

The last thing explored this semester were the harmonics of both the raw audio and the GCI derivatives. The script that processes the GCI derivatives into their harmonic equivalents has yielded some insightful results. The script "**analysisGlottalHarmonics.m**" plots every singer's GCI derivative spectrum for each vowel, then fits and exponential to its harmonics. The fitted exponentials are later used to plot the decay coefficients of the harmonics.

An example of a spectral plot is given in figure 7, but the rest may be seen by running the script. The harmonic decay coefficients are plotted in figure 8. This idea owes itself to Thomas D. Rossing's "The Science of Sound", where he ascribes breathy voice to a dominant first harmonic, indicating that the glottis is not vibrating to produce harmonics. Thus, we expect to see higher decay coefficients for untrained than practiced singers. Unfortunately, the analysis by decay coefficients did not prove to be fruitful. To verify this, the breathy voices were qualitatively verified to be correct by listening to each recording and noting the vocalist's subjective characteristics. The most insightful result of this analysis is that the practiced singers' harmonic peaks were very wide, indicating a very sharp GCI derivative, and the opposite is true for untrained singers. This is a great reality check that the analysis being done is still accurate.

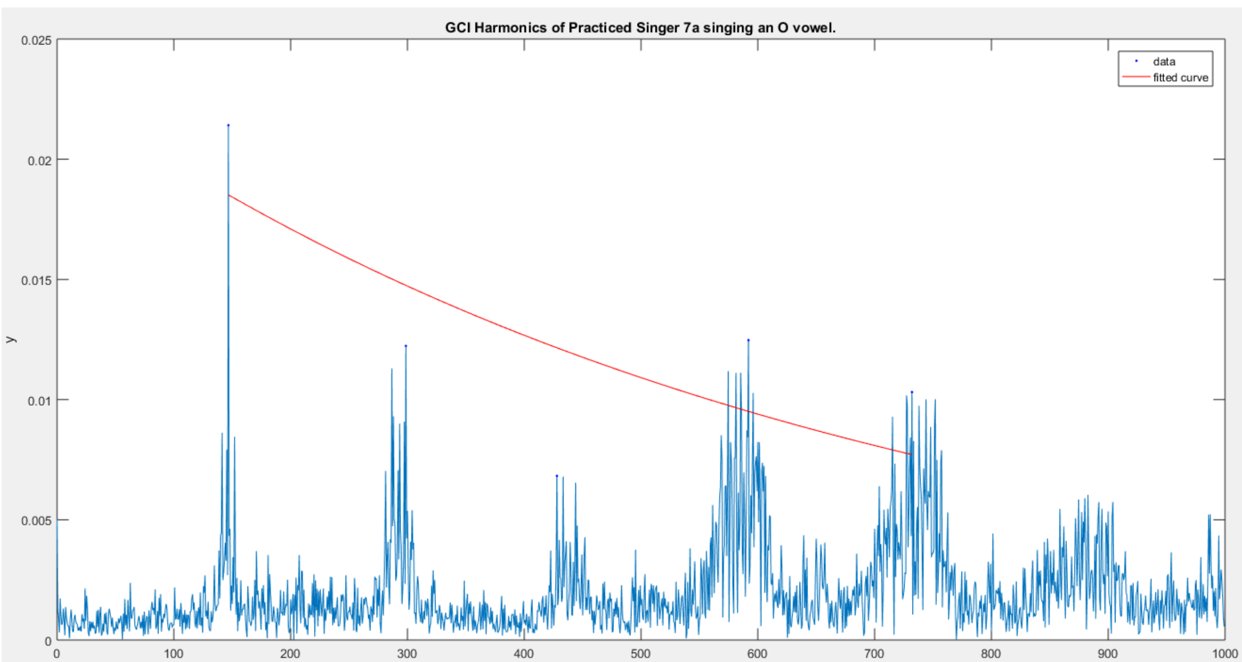

Figure 7. An example of a GCI derivative spectral plot with a fitted exponential.

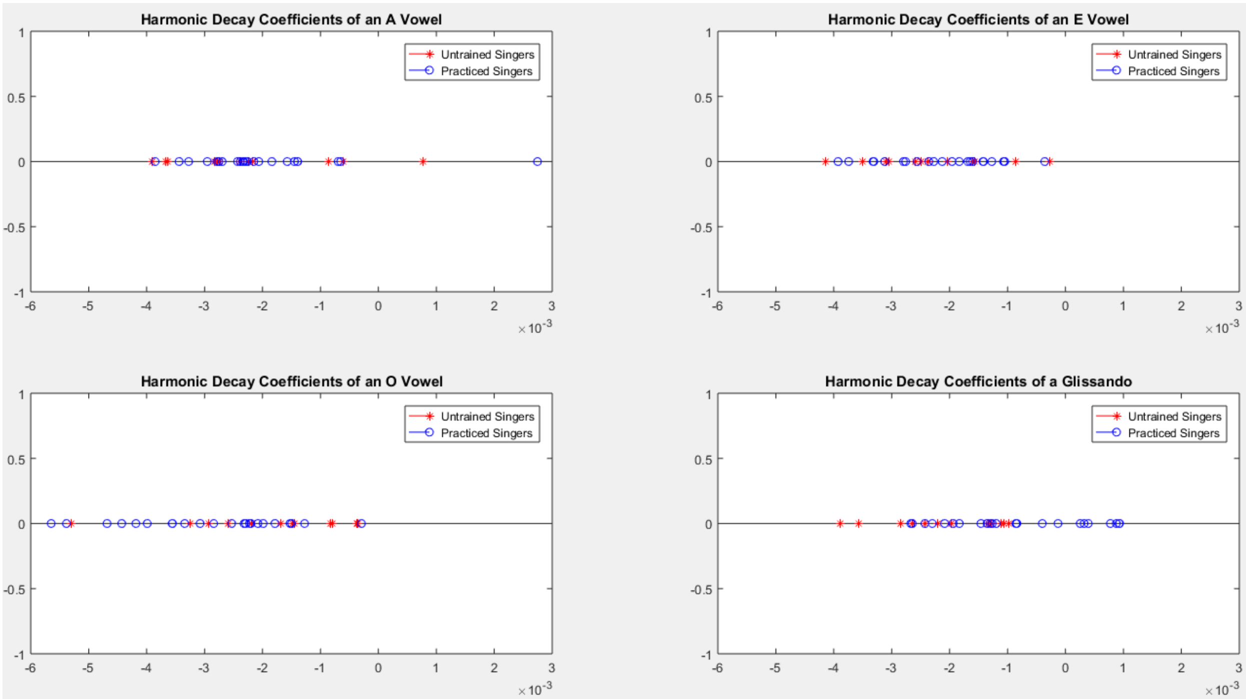

Figure 8. Plot of harmonic decay coefficients for each vowel and the glissando.

In this work, I only wish I had had time to better sort men and women vocalists for analysis and had had time to look at GCI derivative peak widths and GCI derivative peak consistency. I hope to accomplish this in December, so I will send a revised version of this final report in January with that appended. However, other than that, I think this has been a fruitful analysis that has shed light on the quantifiable differences between practiced and untrained singers, and I myself have learned a great deal in the process. Maybe this project continues further, but if it does not, I am very happy having gotten this far.