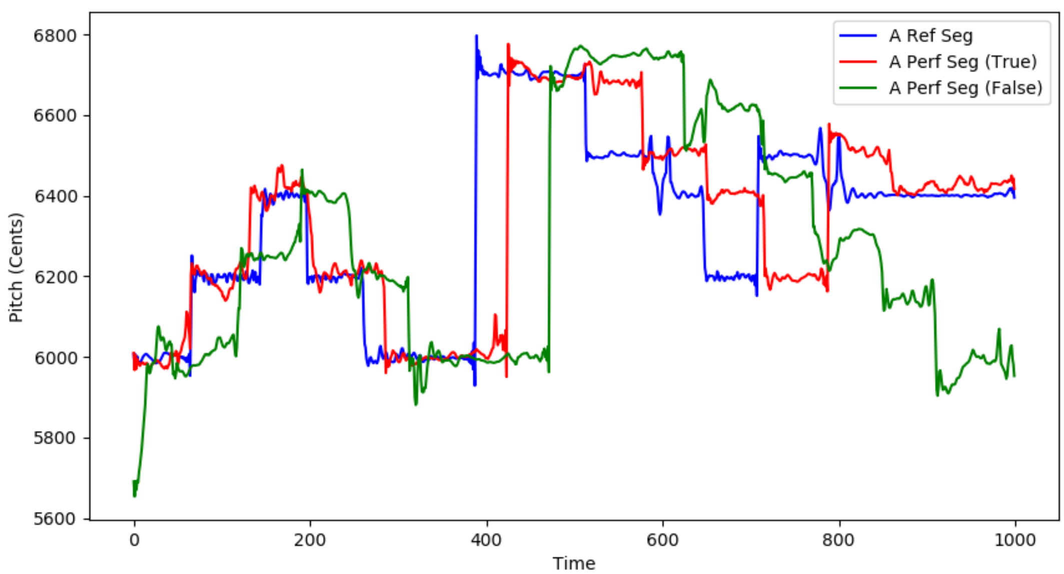
Elena Georgieva
MASTdataset Project

1. After downloading the data and running the Julia files to create the .json files, the first thing I did was plot the data. Here is a pitch track for one of the provided melodies. On this plot, in blue is a piano reference recording with steady and reliable pitch. In red, you can see the pitch track of someone who passed the vocal examination, and in green is the pitch track of someone who failed. The student who failed struggles with pitch especially at the end of the segment and especially lags in timing.



2. Next, I did a statistical correlation: Pearson's r. Pearson's r is a correlation coefficient that measures the linear correlation between two variables. In this case I was comparing the reference segment to a passing performance of the same melody as well as to a corresponding failing performance. Results have been included below. In short, a passing performance on average has a higher r magnitude than a failing performance. The median r value for a passing performance was .94 with the maximum up at .96. For failing performances the mean was .78 with a max of .94. While this is an encouraging metric, both variable pairings have high r values and there is a ceiling effect.

-- Results --
True Min: 0.9391518186720654
True Max: 0.9578295210459282
True Mean: 0.9461610726297641
True Median: 0.9415018781712986
False Min: 0.42567343523820417
False Max: 0.9432589228330768
False Mean: 0.7690261249654489
False Median: 0.7843011443174728

3. Finally, I attempted a classification method. I chose to try support vector machines (SVM), which is a supervised machine learning method used for classification. SVMs are effective when the data have many dimensions, SVMs are versatile, and they are common first algorithm to try when attempting a classification task.

I had 38 .json files and chose to use 30 for training and 8 for testing. The python SVM sklearn function has an argument which is the kernel function. I tried several including a linear and sigmoid kernel, but found a polynomial of degree 3 to be most effective.

I ended up with a precision score of .229. A precision score evaluates the accuracy of positive predictions. I also got a recall score of .367, and this score represents the fraction of positive passing classifications that were correct. This result is poor and actually worse than chance. It is interesting to look at the printed confusion matrix, included with results below. Based on the confusion matrix, a true negative was classified 99 times and a false positive 37 times. These are significantly larger values than the 19 and 11 for the true positive and false positive results, respectively. Looks like this SVM algorithm was classifying more melodies as negative overall. This may be because of a data mismatch, we have many more pitch tracks of students who failed the exam than passed. Also, The original Bozkurt 2017 paper did Dynamic Time Warping before doing the machine learning. Dynamic time warping is an alignment algorithm which may help score a singer while putting less weight on their time and rhythmic performance, which doesn't seem to be as relevant to the task. In the future it would be great to try DTW, balancing the dataset, and other machine learning algorithms.

```
# -- SVM Results --
Confusion Matrix:
[[99 37]
 [19 11]]

              precision    recall  f1-score   support

           0       0.84      0.73      0.78       136
           1       0.23      0.37      0.28        30

    accuracy                           0.66       166
   macro avg       0.53      0.55      0.53       166
weighted avg       0.73      0.66      0.69       166

Precision score: 0.22916666666666666
Recall score: 0.36666666666666664
```

References

1. Bozkurt, B., Baysal, O., Yuret, D. (2017). A Dataset and Baseline System for Singing Voice Assessment, 13th Int. Symposium on Computer Music Multidisciplinary Research, Porto, Sept. 25-28, 2017.

2. Faghih, Behnam; Timoney, Joseph. (2019) Considerations for the Next Generation of Singing Tutor Systems. AES E-Library. March 10, 2019

3. Simon Waloschek, & Aristotelis Hadjakos. (2018). Driftin' Down the Scale: Dynamic Time Warping in the Presence of Pitch Drift and Transpositions. In Proceedings of the 19th International Society for Music Information Retrieval Conference (pp. 630–636). Paris, France: ISMIR.