Hannah Portes
COMS 6998 Advanced Topics in Spoken Language Processing
Homework 3: Feature Analysis

Method
I used z-score normalization method.

Calculation
The calculation is all covered and detailed in the corresponding notebook (.ipynb) file.

To normalize the pitch and intensity arrays produced by Praat, I first grouped all the rows in my data frame by speaker. Using a loop, I selected only the rows corresponding to one speaker at a time, and concatenated all values of pitch and intensity to calculate an overall mean and standard deviation to use in my calculation to normalize the arrays. Following these calculations of the mean and standard deviation of pitch and intensity, I sorted row by row of my filtered data frame (only 1 speaker at a time) and normalized each value in the arrays using the formula

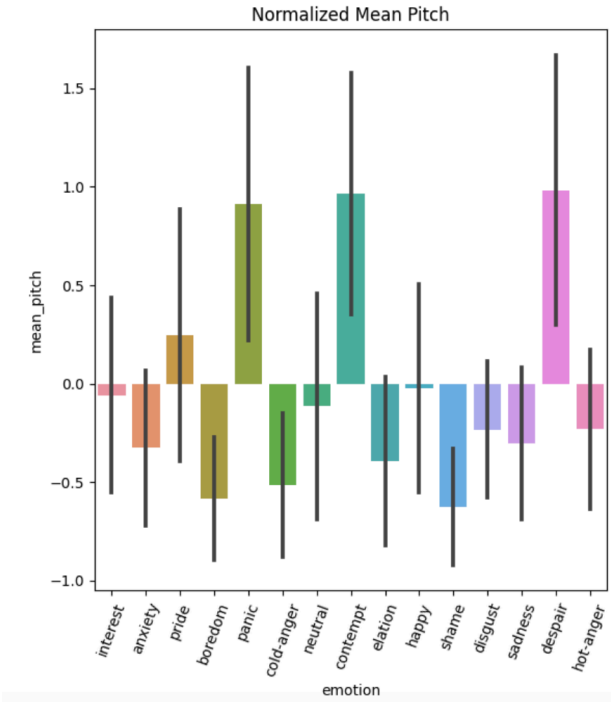
$$(\boldsymbol{x} - \boldsymbol{\mu}_X) / \boldsymbol{\sigma}_X$$

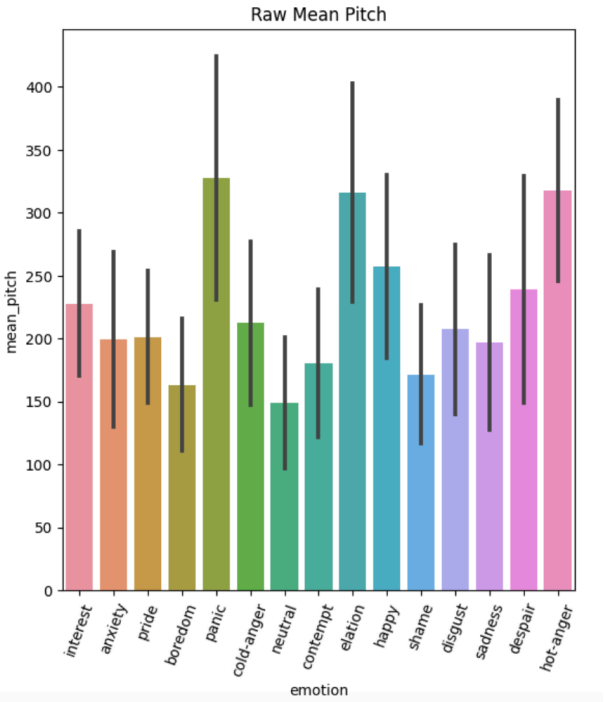
After normalizing the pitch and intensity arrays for each audio file by speaker, I stacked on the resulting normalized values to a new data frame (normalized_data) until all speakers had the intensity and pitch arrays normalized by each individual speaker.

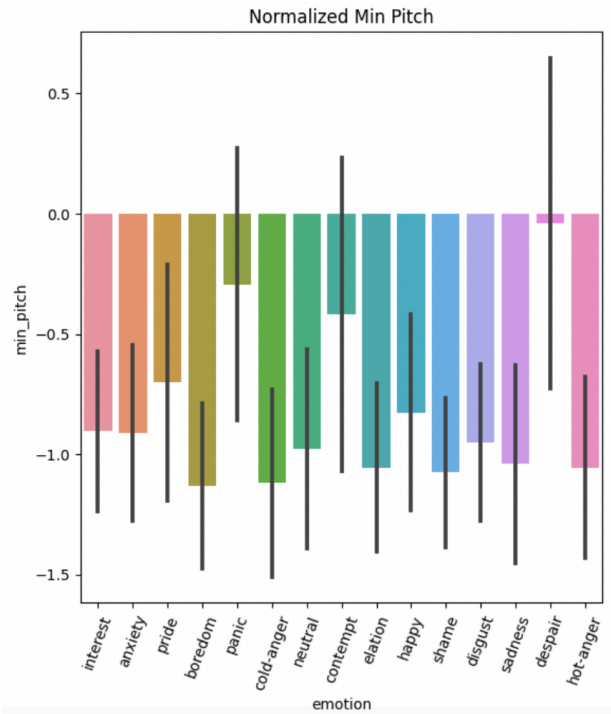
After producing this normalized data frame I added on the mean, max, and min intensity and pitch values calculated row by row from each of the normalized arrays.

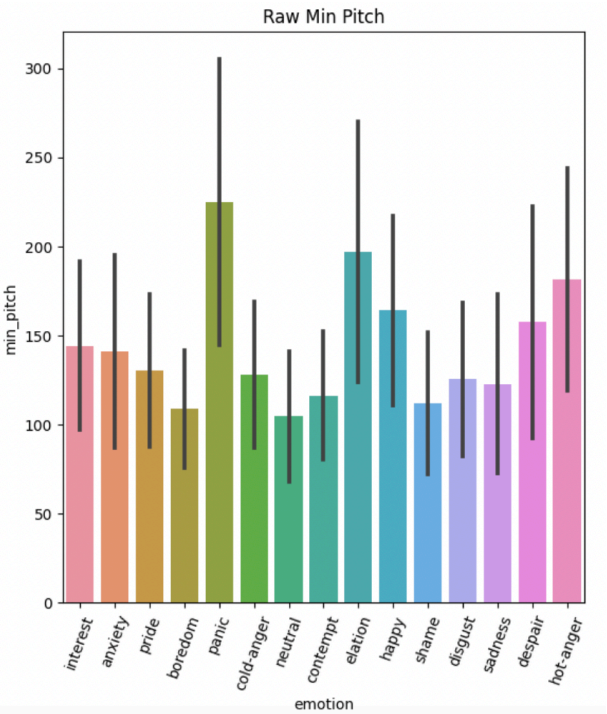
Why this method
I chose this method because the data did not appear to have such extreme outliers that it required clipping. Additionally, z-score can show where the data lies in the entire distribution. While z-score normalization does not ensure all features have the same scale as min-max normalization would, it sufficiently handles outliers. When applicable, I prefer for the data to be squeezed down to a mean of 0 and standard deviation of 1 as z-score normalization achieves.

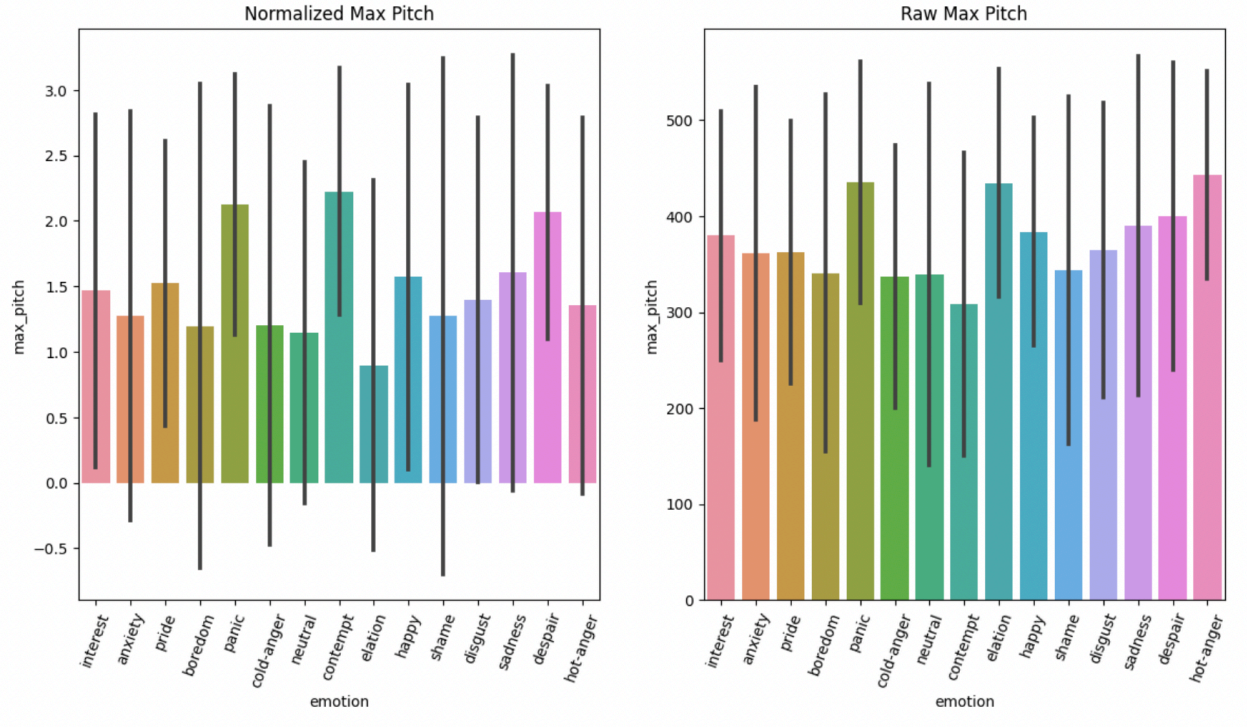
Pitch Plots


Normalized Mean Pitch
mean_pitch
1.5
1.0
0.5
0.0
−0.5
−1.0
interest
anxiety
pride
boredom
panic
cold-anger
neutral
contempt
elation
happy
shame
disgust
sadness
despair
hot-anger
emotion

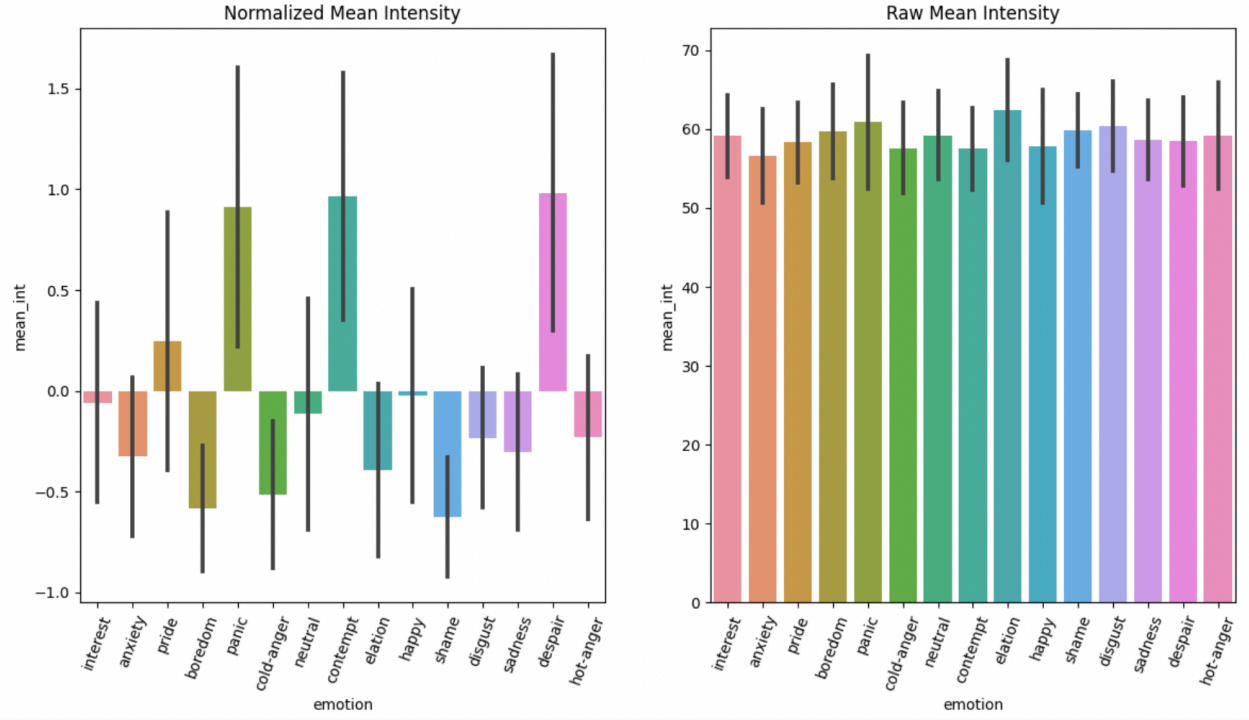

Raw Mean Pitch
mean_pitch
400
350
300
250
200
150
100
50
0
interest
anxiety
pride
boredom
panic
cold-anger
neutral
contempt
elation
happy
shame
disgust
sadness
despair
hot-anger
emotion

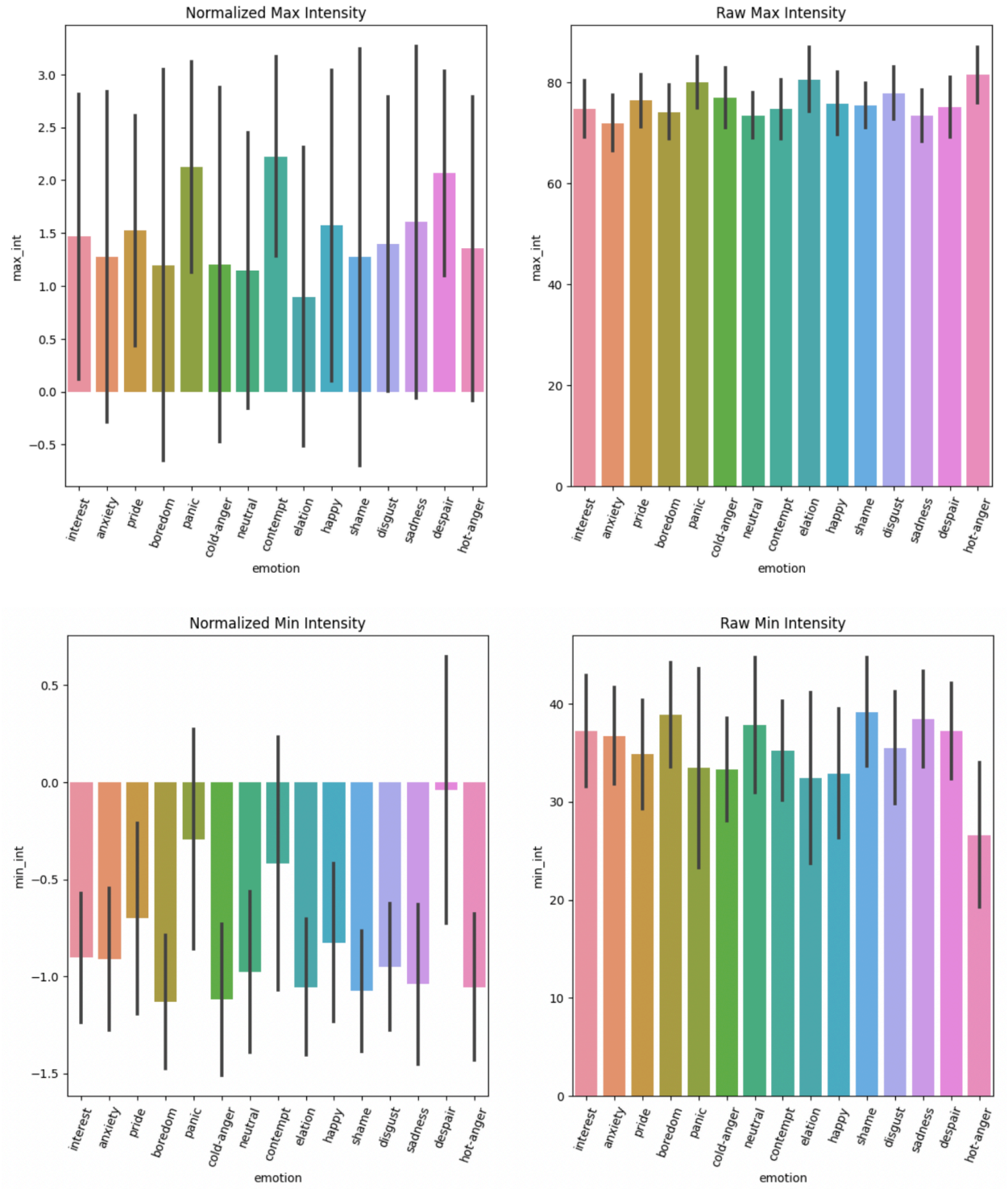

Normalized Min Pitch
min_pitch
0.5
0.0
−0.5
−1.0
−1.5
interest
anxiety
pride
boredom
panic
cold-anger
neutral
contempt
elation
happy
shame
disgust
sadness
despair
hot-anger
emotion


Raw Min Pitch
min_pitch
300
250
200
150
100
50
0
interest
anxiety
pride
boredom
panic
cold-anger
neutral
contempt
elation
happy
shame
disgust
sadness
despair
hot-anger
emotion



Intensity Plots





Observations

1. The first observation I found interesting was that the raw intensity mean, min, and max values are much more similar across all emotions than the same values for raw pitch are. The raw values of max, mean, and min pitch have far more variation as well as the standard deviation of these non-normalized values. Since intensity can be perceived as sound volume, it makes sense that the intensity across all emotions are far closer in raw values than pitch.

2. Panic, contempt, and despair have large ranges in values and standard deviations of normalized mean pitch and normalized mean intensity. Panic, contempt, and despair have significant standard deviations when plotted out. Looking into the emotions that fall into this category makes some sense as the expression of these emotions could vary greatly. I compare these distributions to an emotion such as neutral, which almost universally would be expressed fairly flat with little change in extracted features.

3. Interest, happy, and neutral have low variations in mean pitch and mean intensity values. I interpret this to mean that the values for these emotions are more distinct and recognizable to a model than other emotions. These emotions that have non-distinct and small distribution of datasets could essentially blend in to other emotion classes easily and thus result in poor model performance for these classes. This could also be a result from the proportion of samples in these classes, but looking into the data this is not the case as interest and happy have an average proportion of samples (only neutral is significantly lower at 70 samples).

4. These plots highlight that raw values really do not inform us much about features extracted from these audio recordings and normalization is essential to interpreting results. It would be interesting to explore different normalization techniques to see if or how these impact the resulting graphs. I believe they would not have a drastic effect as all normalization methods should scale the features to a more standard scale between emotions. This would still be interesting to explore and verify, perhaps outliers have more of an impact on this dataset than I interpreted and z-score normalization is not the optimal method.

5. Despair, despite having one of the largest ranges for normalized mean and max intensity and pitch, has the smallest distribution of min pitch and intensity values. This implies that the minimum values for pitch and intensity are fairly consistent across speakers once normalized as opposed to the other emotions that have significantly larger ranges for normalized min pitch and intensity.

6. I found it interesting that hot-anger also has quite a small standard deviation in mean pitch and intensity, but also a lower normalized value for both. I would expect quite a variation in hot-anger and I am interested how this was defined and explained to participants. Hot-anger I would consider anything from yelling to a very deep stern monotone but the smaller standard deviation implies different.