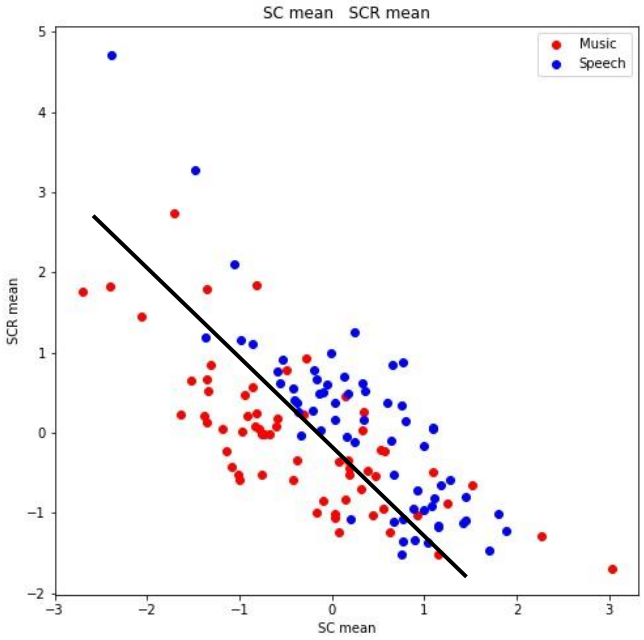
Feature Extraction: Analysis

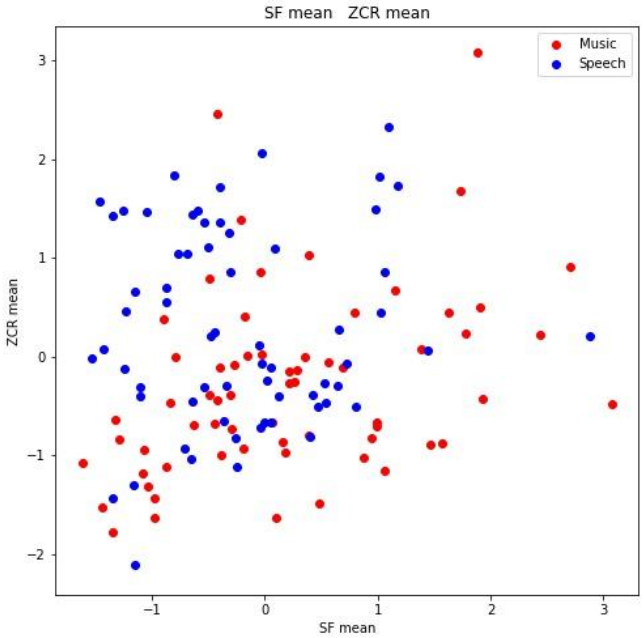
Here's a summary of inferences from each of the plots-

1. SC mean vs SCR mean: A combination of SC mean and SCR mean can be a decent classifier (along the black line). It's clear that for both, SC mean and SCR mean, speech and music can correspond to high and low values but it's interesting to see that some classification can be done on the basis of the combination. Most of the points on the left side of the classifier correspond to music files. It can be said that if the weighted sum of SCR mean and SC mean is less than a constant value (equation of a straight line), it is more likely to be music.



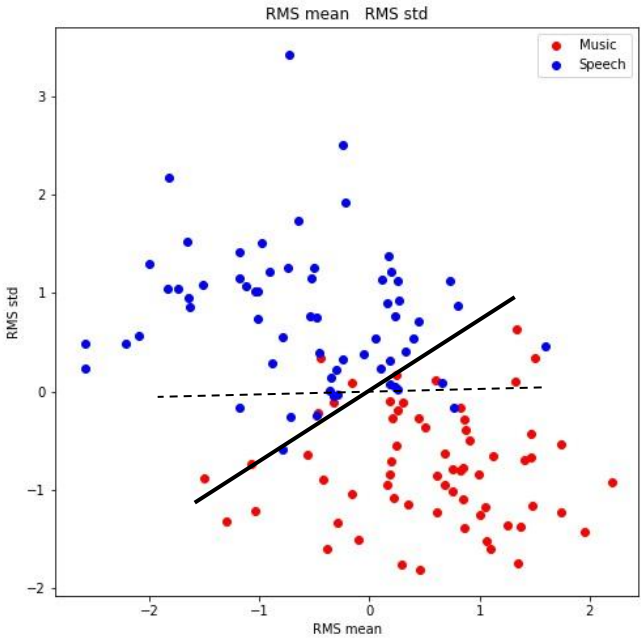
Plot 1: SC mean vs SCR mean

2. SF mean vs ZCR mean: It can be observed that there is no visible pattern for distinction. One could argue that music has higher SF mean because barely any speech signal has SF mean of more than 1, while there are a handful of music signals with SF mean > 1. But majority of datapoints overlap with speech in -1< SF < 1 range.



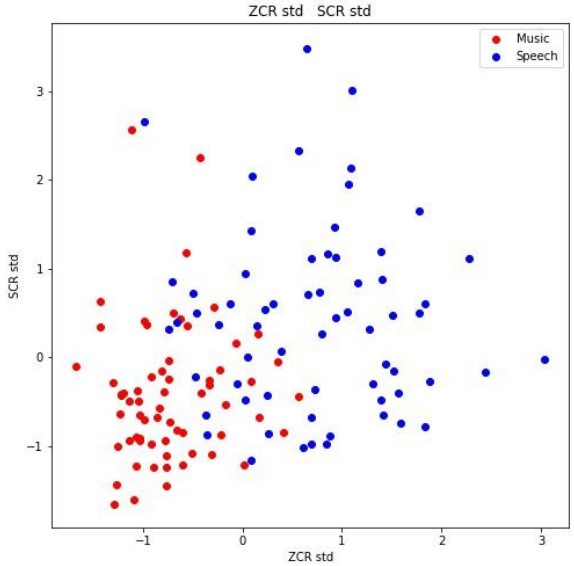
Plot 2: SF mean vs ZCR mean

3. RMS mean vs RMS std: It can be said that music audio has higher RMS std as compared to speech audio (dotted line, just based on RMS std). But if we use the solid black line as a classifier, then we can get better accuracy. Hence, a combination of RMS mean and RMS std is another good set of features to use. There is no clear difference observed across RMS mean as such.



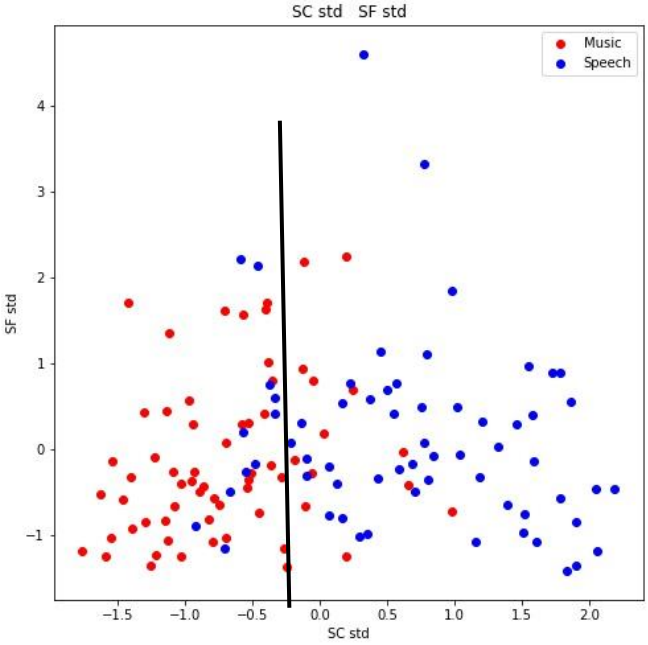
Plot 3: RMS mean vs RMS std

4. ZCR std vs SCR std: Differentiation between music and speech along ZCR std is much more than that along SCR std. Music tends to have lower ZCR std as compared to speech. There is some overlap but ZCR std can be used as a classifying feature. In case of SCR std, majority of music and speech files lie between -1 and 1. Due to this strong overlap, it doesn't seem to be relevant for classification



Plot 4: ZCR std vs SCR std

5. SC std vs SF std: Distinction along SC std can be observed. A potential classifier can be a line parallel to Y-axis, somewhere between x = -0.5 to 0. No difference in SF std is observed between music and speech data.



Plot 5: SC std vs SF std

Overall learnings-

1. RMS is a good feature to extract. Combination of mean and std (Plot 3) classifies the two categories very cleanly
2. Spectral centroid can be a good feature. Both, Std dev by itself (Plot 5) and SC mean combined with SCR mean (Plot 1) can be used for classifying
3. ZCR can be another important feature. We can see that ZCR std is a good differentiator (Plot 4). However, ZCR mean has only little potential of differentiating.
4. Looking at plot 2 and 5, it can be said that Spectral Flux is probably not a good feature to differentiate between speech and music.
5. SCR isn't that important either. Along with SC mean, SCR mean could classify, but there isn't much differentiation in SCR mean and std for music and speech.