Identifying Heart Rate: A Data Analysis Project

In our world today, technology is developing at a rapid rate and presents itself in many forms as a part of daily lives. As technology continually and increasingly integrates into human lives, professionals and experts seek to utilize technology for purposes of human health. Specifically, one application that has been pursued in research is the potential to measure heart rate via motion sensors such as accelerometers and gyroscopes. In this data analysis project, I built algorithms to extract heart rate from motion sensor data and examined the accuracy between the provided ground truths of heart rate from data collection and the predicted heart rates.

Outlined by three key stages, my project consists of preparation, visualization, and analysis. To prepare the data for subsequent steps, the raw data was modified minorly. The beginning and ending values of the raw data were removed due to unusually high/low values that derail the analysis. These abnormal values are due to data collection device flaws.

Prior to analyzing, I visualize the data by plotting them. Examining the graphs formed by unanalyzed data (and later graphs based on analyzed data), I compared the graphs to draw inferences that were otherwise unknown. For instance, a few of the raw datasets present almost completely straight lines when plotted. From this, I gathered that something went wrong during the process of data collection and eliminated the possibility that it was an error in my program since other raw data graphs plotted fine.

Lastly, the majority of this project consists of data analysis. Following preparation and visualization, I implemented a four-step algorithm to move towards extracting heart rates from the data. First, I implemented moving averages in the raw data to smooth out fluctuations and thus reveal underlying trends better. Next, I implemented a band-pass Butterworth filter to eliminate unwanted frequencies that are too high/low. With the filtered data, I then calculated the L2 Norm to combine the components of the signal. The last step of the four steps is filtering the data with a band-pass Butterworth filter again, but this time on the L2 Norm calculated in the previous step and also with different parameters. Following this, the filtered L2 Norm data was transformed from the time domain into the frequency domain using Fast Fourier Transform. Now, the heart rate was ready to be extracted. To find the indicated heart rate, I identified the peaks in the transformed L2 Norm data and found their corresponding frequency. From here, it is simply multiplying the frequency by 60 in order to calculate the indicated heart rate. However, the highest peak is not always accurate in indicating the heart rate. Hence, I assessed the three highest peaks then calculated the error between the corresponding frequencies and the ground truth (when given). Then, the one with the least error presents itself as the predicted heart rate. After completing all of the above steps on each dataset, I compiled and calculated all of the errors in heart rate prediction and grouped them by sensor. In a spreadsheet, the level of error that each sensor is associated with is indicated by their absolute mean error, standard deviation of the absolute error, and root mean squared error. Examining my table of compiled errors, it appears that no particular sensor offers substantially less error. However, it does appear that the accelerometer offers slightly less error, compared to the rest of the sensors.