# CMPT 353 Final Project Report

## Topic: Sensors, Noise, and Walking

Alfred Rodillo (arodillo@sfu.ca)
Janit Kumar (janitk@sfu.ca)
Huy Nguyen (hnn3@sfu.ca)

# Introduction

In this project, we are interested in accelerometer data collected from a person walking. This report will discuss how the data can be used to determine individual attributes along with calculating distance, velocity, etc. The report also contains information about how the data was collected, and how it was filtered and transformed using data science techniques. Moreover, how the data is analyzed to obtain and visualize the results through graphs and plots is also added in the report.

# Data Collection

The data was collected by each group member using a mobile app, Physics Toolbox Sensor Suite. With this app, we obtained linear accelerometer data collected by the phone's built-in sensors. The data was stored by the app in a .csv file format to be processed later. Our data was collected by having the person walk in a straight line for roughly sixty seconds, with the time taken to start and stop the recording added to that total. For comparisons, the data was collected with the phone in three different locations: held in hand, placed in a pocket, and placed at the ankle. Walking data was collected with the phone in each location at least twice from each group member. A data set is also included that includes real step counts in order to verify step frequency in a later stage.

# Data Processing

## Filtering

Since the data had a lot of noise due to phone sensors it was filtered using a low-pass Butterworth filter of order 3 in an attempt to isolate only the frequencies we are interested in. The data was collected by a walking person, so we expect the step frequency to be fairly low. In the paper referenced in this course's project topic page [1], they used a filter with a cutoff frequency of 8 Hz according to the Fast Fourier Transform (FFT) of their data. Figure 1 shows the frequency spectrum of one group member's walking data, collected with the phone in hand.

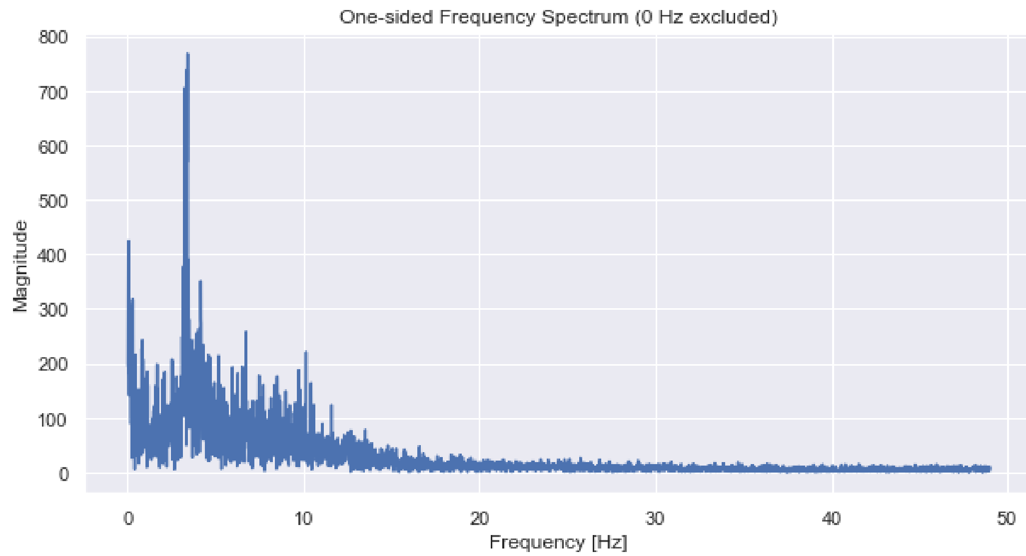


Figure 1. Frequency spectrum of total acceleration recorded with the phone held in hand.

Similar distribution features were also found in other members' data, and with the data collected with the phone in a different location. The frequencies with the largest magnitudes are mostly situated on the lower end of the frequency range, so our cutoff frequency was set at 5 Hz. Figure 2 shows the frequency spectrum of the same data in Figure 1 after applying the filter.

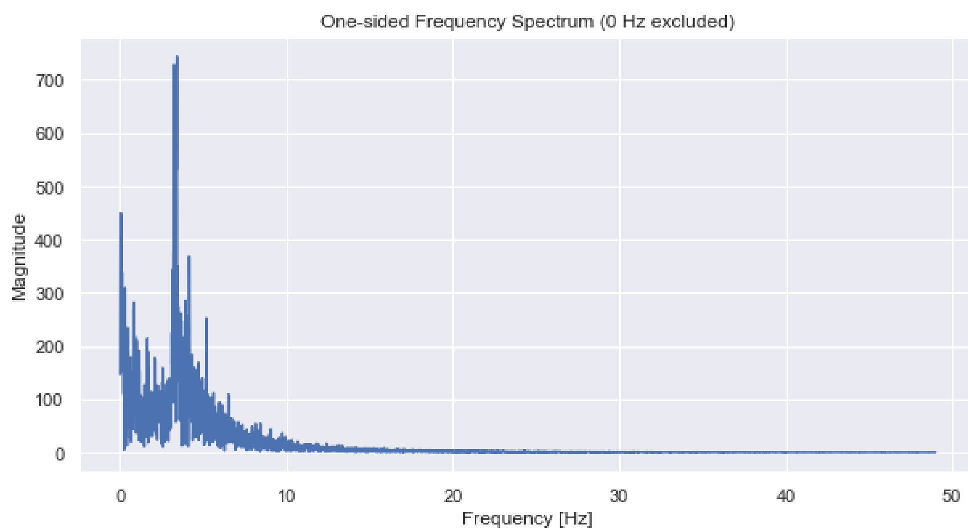


Figure 2. Frequency spectrum of the filtered data set used in Figure 1.

## Transformation

The data needed to be transformed according to the requirements of the functions and other team members. For some data sets, the time column had to be changed from DateTime format to milliseconds because that is easier to read and understandable. In another data set, the column headers in the .csv file were different from those provided by other group members and needed to be formatted for convenience.

# Analysis

## Step Frequency

From the filtered acceleration data, we attempted to determine an individual's step frequency, measured in Hz (or steps per second). Since the data was recorded with approximately sixty seconds of walking, the individual would have reached a steady pace, and that should occupy the majority of the observed signal. However, the frequency spectrum of the signals showed a large spike at 0 Hz, pointing to that as the dominant frequency in the total acceleration. That result is not meaningful given the context, so the 0 Hz value has been excluded from the rest of the analysis, as shown in Figures 1 and 2.

The ten frequencies with the largest magnitudes are then taken as "candidate frequencies", from which a single value for step frequency will be calculated. There were three methods to determine the step frequency. Using the data set with real step counts, they can be compared for correctness at different phone recording positions. The first method is to simply take the frequency with the largest magnitude. The second is to take the mean of all candidate frequencies. Lastly, the third method is to take the mean of candidate frequencies whose magnitudes were at least half of the largest within that set.

## Speed and Distance

We extracted the average walking speed, in meters per second, and total distance walked, in meters from the filtered dataset. Given that each person walks with a steady pace on a mostly straight path, the output for the average speed looks to be accurate, with a few outliers in our 'hand' datasets. The speed of the phone held in hand ranges from 2.14 m/s to 38.54m/s, while the speed of the phone tucked in pocket looks to be the most consistent across all datasets, ranging from 1.30m/s to 5.01m/s. Similarly, since distance is directly proportional to speed, the total distance walked when the phone is in the pocket also reflects accurately in accordance with our input. The largest value for the distance walked in our filtered dataset is 2312.85 meters, which is impossible for a walking pace in a span of a minute.

# Results

Figure 3 shows the comparisons of the different methods to determine step frequency, using the data set with real step count.

| | filename | real_steps | freq1 | steps1 | freq2 | steps2 | freq3 | steps3 |
|---|---|---|---|---|---|---|---|---|
| **0** | ankle1_100.csv | 100 | 0.011661 | 1 | 1.315357 | 113 | 0.876904 | 75 |
| **1** | ankle2_108.csv | 108 | 0.011696 | 1 | 1.632796 | 139 | 1.148182 | 98 |
| **2** | ankle3_112.csv | 112 | 0.011057 | 1 | 1.479388 | 134 | 1.159112 | 105 |
| **3** | ankle4_112.csv | 112 | 0.011193 | 1 | 1.289480 | 115 | 0.870847 | 78 |
| **4** | hand1_114.csv | 114 | 1.758506 | 126 | 0.815053 | 58 | 1.758506 | 126 |
| **5** | hand2_114.csv | 114 | 1.739272 | 122 | 1.069225 | 75 | 1.739272 | 122 |
| **6** | hand3_112.csv | 112 | 1.743913 | 124 | 0.983060 | 70 | 1.736881 | 123 |
| **7** | hand4_111.csv | 111 | 1.713106 | 130 | 1.043677 | 79 | 1.726283 | 131 |
| **8** | pocket1_114.csv | 114 | 0.010644 | 1 | 1.167701 | 110 | 0.305142 | 29 |
| **9** | pocket2_117.csv | 117 | 0.012046 | 1 | 1.874362 | 155 | 0.368608 | 31 |
| **10** | pocket3_117.csv | 117 | 0.012424 | 1 | 2.226293 | 179 | 1.276072 | 103 |
| **11** | pocket4_114.csv | 114 | 0.012461 | 1 | 1.262325 | 101 | 0.720260 | 58 |

Figure 3. Table containing the predicted step frequency and step count with different methods.

The naive first method was unable to get close to a reasonable estimation for most of the data. The low frequencies may have dominated the signal too much for ankle and pocket data due to additional movement needed to start and stop the phone at those positions. However, it performed well with the data collected with the phone in hand, seemingly able to capture the periodicity from the arms swinging and get close to the real step count. The results with the hand data were also fairly consistent across the four recordings.

The second method performed better on the ankle and pocket data sets, but worse with the hand. The ankle and pocket data sets may have had more reasonable frequencies dominate the candidate set, so examining all candidate frequencies led to better results. On the other hand, there may have been more noise than first realized with the hand data set, contributing to its worse results. Within different positions, the frequencies are not as consistent as they were with the first method.

The third method showed varying results, with the hand data set once again showing more accurate results. Somewhat similar to the second method, we attempt to look at more than just one frequency from the candidate set, but are limited to those with a high enough magnitude. The ankle and pocket data do not show any form of consistency, as well as poor estimations. The hand data set once again showed great consistency, similar to the first method.

Figure 4 shows a summary that includes estimated step frequencies for all collected data, as well as speed, distance, name of the group member, and the position of the phone.

| | filename | freq1 | freq2 | freq3 | speed | distance | name | phone_position |
|---|---|---|---|---|---|---|---|---|
| 0 | ankle4_112.csv | 0.011193 | 1.289480 | 0.870847 | 1.560468 | 93.556254 | ALFRED | ankle |
| 1 | pocket4_114.csv | 0.012461 | 1.262325 | 0.720260 | 3.205848 | 192.380535 | ALFRED | pocket |
| 2 | ankle2_108.csv | 0.011696 | 1.632796 | 1.148182 | 2.316984 | 138.967825 | ALFRED | ankle |
| 3 | hand4_111.csv | 1.713106 | 1.043677 | 1.726283 | 20.879607 | 1253.758568 | ALFRED | hand |
| 4 | ankle3_112.csv | 0.011057 | 1.479388 | 1.159112 | 1.323693 | 79.471425 | ALFRED | ankle |
| 5 | hand1_114.csv | 1.758506 | 0.815053 | 1.758506 | 31.033426 | 1862.115273 | ALFRED | hand |
| 6 | pocket2_117.csv | 0.012046 | 1.874362 | 0.368608 | 4.846752 | 290.778921 | ALFRED | pocket |
| 7 | pocket3_117.csv | 0.012424 | 2.226293 | 1.276072 | 4.443051 | 266.607336 | ALFRED | pocket |
| 8 | hand2_114.csv | 1.739272 | 1.069225 | 1.739272 | 30.335498 | 1820.258343 | ALFRED | hand |
| 9 | ankle1_100.csv | 0.011661 | 1.315357 | 0.876904 | 1.927474 | 115.694341 | ALFRED | ankle |
| 10 | pocket1_114.csv | 0.010644 | 1.167701 | 0.305142 | 4.935996 | 296.143698 | ALFRED | pocket |
| 11 | hand3_112.csv | 1.743913 | 0.983060 | 1.736881 | 23.866346 | 1432.049286 | ALFRED | hand |
| 12 | ankle2.csv | 0.861538 | 1.461026 | 1.286325 | 2.488046 | 149.286344 | HUY | ankle |
| 13 | ankle1.csv | 0.857068 | 1.802402 | 1.509464 | 14.322956 | 860.068918 | HUY | ankle |
| 14 | pocket2.csv | 3.323073 | 3.114615 | 3.277528 | 1.386040 | 83.166085 | HUY | pocket |
| 15 | hand1.csv | 3.412018 | 3.334292 | 3.334292 | 2.140042 | 129.721822 | HUY | hand |
| 16 | pocket1.csv | 3.578763 | 3.033365 | 3.105556 | 1.302673 | 78.150213 | HUY | pocket |
| 17 | hand2.csv | 3.276957 | 2.937388 | 3.265084 | 7.085705 | 425.331865 | HUY | hand |
| 18 | Pocket1.1.csv | 1.782387 | 2.451563 | 2.896378 | 5.012226 | 300.746954 | JANIT | pocket |
| 19 | Hand1.1.csv | 0.014833 | 1.090207 | 0.652641 | 38.544729 | 2312.853784 | JANIT | hand |
| 20 | Pocket1.2.csv | 0.048507 | 0.515789 | 0.515789 | 1.939987 | 116.417098 | JANIT | pocket |
| 21 | Hand1.2.csv | 1.919081 | 1.173653 | 2.145088 | 10.763915 | 645.894625 | JANIT | hand |
| 22 | Ankle1.1.csv | 0.838348 | 1.874720 | 1.058776 | 2.094173 | 125.677246 | JANIT | ankle |
| 23 | Ankle1.2.csv | 0.893529 | 1.970088 | 0.751814 | 2.691984 | 161.528664 | JANIT | ankle |

Figure 4. Table showing a summary of all data collected by group members.

Based on the data summary, we trained a few machine-learning models to predict step frequency in relation to each person. Figure 5 shows the prediction scores for trained models.

```
Bayesian classifier:      0.667    0.667
kNN classifier:           0.833    0.667
Rand forest classifier:   0.944    0.667
```

Figure 5. Scores for various models trained to predict the person based on step frequency.

# Conclusion

The results we obtained were not only dependent on the phone position, but also on what we were trying to look for. When analyzing step frequency, the data set collected with the phone in hand showed the greatest consistency and came close to estimating the step count. However, for speed, the results with the pocket data set showed greater consistency compared to those with the ankle and hand data sets.

Further improvements to our data collection and analysis methods would help to obtain better results. The duration of recording could be limited to just a time period when a person has already achieved a steady walking pace. Alternatively, this could have been done in the analysis stage, by simply taking a slice of the data to analyze for frequency. A better sensor could also potentially minimize the effects of noise, compared to using what is available on our mobile devices, which may have had varying levels of effectiveness.

# References

[1] M. Yousefian, “Design and implementation of a smartphone application for estimating foot clearance during walking,” Summit Research Repository, 05-Apr-2017. [Online]. Available: https://summit.sfu.ca/item/17204.

# Project Overview

Alfred Rodillo

- Worked in a small team to analyze accelerometer data collected from a smartphone
- Collected data using personal smartphone device
- Applied filtering techniques to clean up data by examining initial results and selecting appropriate parameters
- Determined step frequency using different methods and criteria
- Compared correctness of estimated results with actual recorded values
- Contributed to preparation of report summarizing project results

Janit Kumar

- Contributed to a small team of 3 to work with accelerometer data and analyzed it using data science techniques.
- The data was collected using a smartphone, there were multiple datasets and different positions in which the data was collected, three main positions were ankle, pocket, and hand.
- Used the Butterworth filter to filter the unwanted noise and Fourier transform to get the step frequency, also training some ML classifiers and plotting graphs for better visualization of the data.
- Got better results after filtering and transforming the data and score predictions with ML classifiers.
- Total acceleration was calculated which gives overall acceleration for all three axes (x, y and z).

Huy Nguyen

- Collected accelerometer data from a smartphone to be used for analysis
- Filtered the data and applied a Butterworth filter and a Fourier transform to reduce noise and retrieve the step frequency
- Calculated the average velocity and total distance walked for each dataset
- Worked on the analysis of the velocity and total distance walked, and the machine learning models part of the report
- Computed the prediction scores for the machine learning models