CMPT 318 Project

Sensors, Noise, and Walking

Group Member:
ZHIZHOU JIANG, zhizhouj@sfu.ca
MANQING ZHU, manqingz@sfu.ca
ZHAOYANG LI, zla143@sfu.ca

Introduction

1. background

We ourselves are generating data every moment no matter you are aware or not. Moreover, most of us are able to collect those data whenever and wherever with a single cell phone. Right when we are walking, ours cell phone are recording hundreds of thousands of data of our gait through its sensors such as accelerometer and gyroscope. It would be a perfect source of data with a little assistance of some App. Our goal is to make the most of the data we get and get as much as interesting information out of it.

2. Project objectives

a) Collect and extract a group of the sensors data while walking
b) Clean the data so that it could be analysed
c) Analyse the cleaned data and get information with regard to the movement or the person

3. Project questions

a) What is the walking pace (steps/minute) of this person given the walking data?
b) What is the walking speed (m/s) of this person given the walking data? How accurate can we get?
c) Whether the gait of this person is symmetrical given the data of left foot and right foot?
d) How does those attributes vary among different people?

Data: collect

1.equipment

Sensor hardware:

iPhone X, iOS 11.1.2

iPhone 7 Plus, iOS 11.1.2

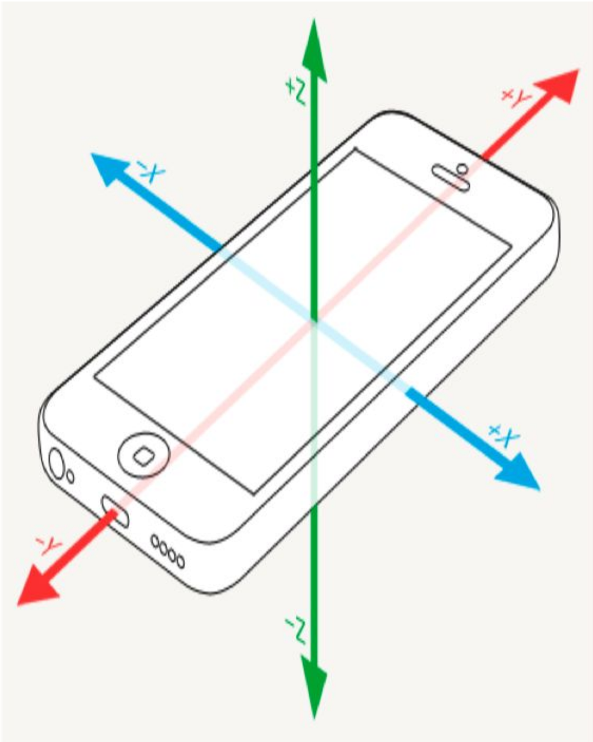
Software:

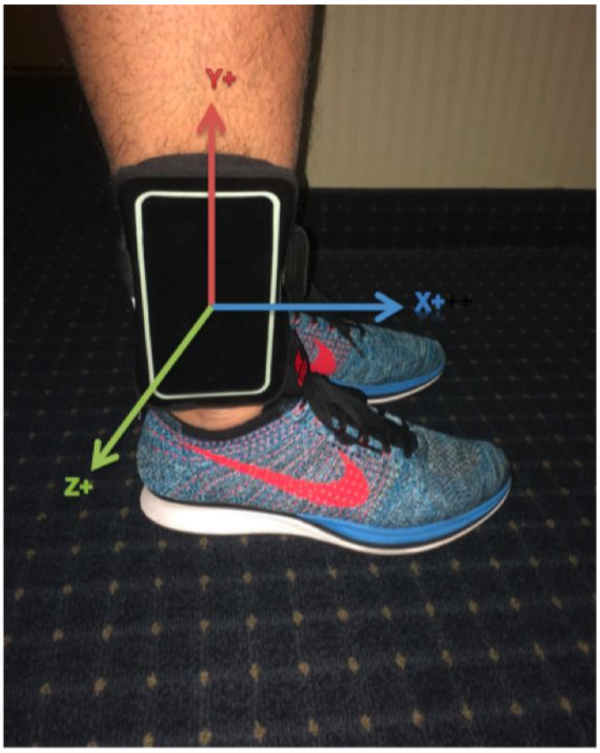
SensorLog version 2.0

Website: http://sensorlog.berndthomas.net/

We collect the data from Accelerometer and Gyro. The data points are sample by 30hz. SensorLog App read the data from Apple Core Motion API, and provided nice and clean csv file we can clean and analyze the walk data.

2.Measurement methods





iPhone Sensor Orientation **Gravity vector of the test**

We choose a straight and clean sidewalk with length 22m, and set two phones on the tester both feet (above the ankle and same side), then let the tester walk straight on the sidewalk. The reason we did this is we want to get the walking data with minimal noise and try the best to eliminate irrelevant variables, and get both feet data at the same time to reduce test error.

Data: clean

1. Extract-Transform-Load (ETL)

For the benefits of the further data processing, we first need to carry out the process of Extract-Transform-Load (ETL) to convert data into a better format.

The data we collected is stored in CSV files. We used the pd.read_csv to read them into DataFrame. The original acceleration data is measured as “nG”, where G is the local gravitational force. Thus, it’s necessary to scale these data back to the unit of m/s^2 before using them to do some calculation. We also used the pd.to_datetime to convert the timestamp data.

Due to the method we used to collect data, only part of data in the original files is produced during walking. To extract the valid data, we plotted the original data and select a time slice to get usable data by observation.

2. Noise Filtering

Naturally, there is a lot of noise in the data we collected from the phone sensors. To remove the noise from signals, we applied the butterworth filter to both accelerometer data and the gyroscope data. The filtering cut-off frequency is chosen taking the result of Fourier Transformation into consideration. We aimed to get the smooth enough signals without any change of their critical properties. For the accelerometer data for three axis, the low-pass filter is used to remove high-frequency noise. For the gyroscope data for three axis, the band-pass filter is applied to remove high-frequency and constant noise at the same time. The choice of cut-off frequencies vary among different experiment entry.

Data: analyze

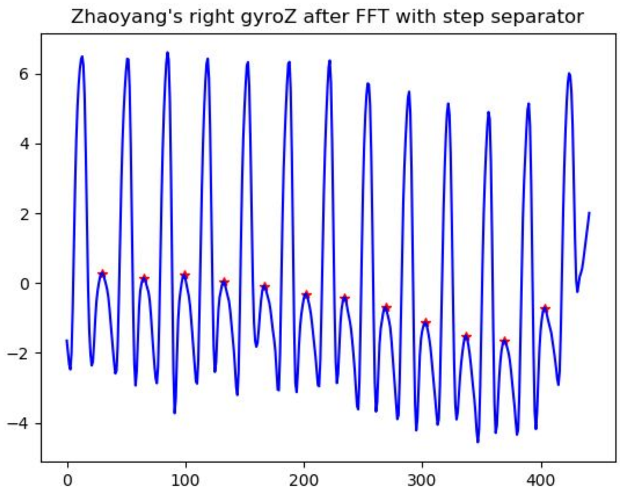
After we got the clean data, we could use them to do some calculations and analysis.
1. Frequency
After cleaning the data, what we got was signals consisting of a series of repeating pattern. It's reasonable to use Fourier Transformation to get the frequency of the pattern in the signals.
According to the paper of Miss Maria Yousefian, the frequency of gyroscope in the Z-axis can be regarded as the frequency of steps.

We got the frequency with numpy's implementation of Fast Fourier Transformation. The function np.fft.fft can convert a list of N complex number to a list of N complex numbers. And then, we can use np.fft.fftfreq to get the corresponding frequencies of the array got from the Fourier Transformation. The step frequency we want is exactly the one associated with the entry which has the maximum absolute value in the array got from the Fourier Transformation.

We have plotted the array returned from FFT and can find that there is a lot of small fluctuations in the curve. If the signal is totally noisy, then there will be a flat line on the figure. Except from the gyroscope data in the Z-axis, we also tried to get the frequency of other data. In the most case, we can get the same frequency or similar frequencies from different sort of data. It suggests that our data measured by the phone sensors is rather accurate. In the process of our experiment, we can such result that the frequency of accelerometer data in the Y-axis is 0 for most of sampling data. It may result from the reality that the measured value of the accelerometer data in the Y-axis is usually very small, thereby difficult to get its frequency.

2. Separate each step
In order to calculate the velocity and displacement of each step, we need to first separate each step. A step begins with so called "mid stance"[1]. It is when the foot is completely on the

ground and the leg is upright. At this point, the absolute value of angular velocity in the Z direction get by gyroscope(gyroZ) is minimized. It is like the middle point of letter "W" as shown by the red star in the figure below.



To find this point, we first separated the data into slices by the frequency we got. Within each step frame, it was each to find the peak. Once we found the peak, which was like the left and right endpoint of letter "W", we went all the way down along the edge to the two bottoms. Then we just need to find the highest point between the two bottoms.

3. Velocity

The idea of calculating velocity is to integrate acceleration with drift correction. We did this step by step.

We Assumed that at mid stance the cell phone is vertical so Vy is 0 and it only have Vx due to rotation around ankle. We measured and recorded the distance between the center of cellphone and the ankle. Multiplying it by gyroZ, we got the initial value of Vx of each step, as well as end value. Then we integrated acceleration and get Vx and Vy at each point in each step cycle.

The velocity calculated this way need to be corrected because at the end of each step the calculated velocity appears to drift away from the theoretical velocity we obtained by angular velocity. So drift correction was necessary. For each point inside a step cycle We did this by the following equation:

$$Vcorrected(t) = V(t) + deltaV/deltaT * t$$

"deltaV" is the difference between theoretical V and calculated V. "deltaT" is the duration of the step, and "t" is the time starting from the beginning of a step cycle.

4. Displacement and walking speed

Once we had velocity, we could apply the same method to get the displacement. In X direction which is the forward direction, the displacement of each step we got reflects the step length. Thus we could easily get the walking speed by sum of step length divided by sum of time interval.

In Y direction which is the upward direction, the displacement should be 0 in the beginning and ending of each step. So we could also apply drift correction on this. We could find the maximum displacement in Y direction during a step which reflects the maximum foot clearance (same thing as height).

5. Stats tests
We did some stats tests on the step length and step duration time of the data get from left and right feet to see whether the gait is symmetric or not. We did normal test, t test and mann whitney u test. The results is the next section.

Results

Dataset 1:
Tester: Zhaoyang Li, male, 20, no injury

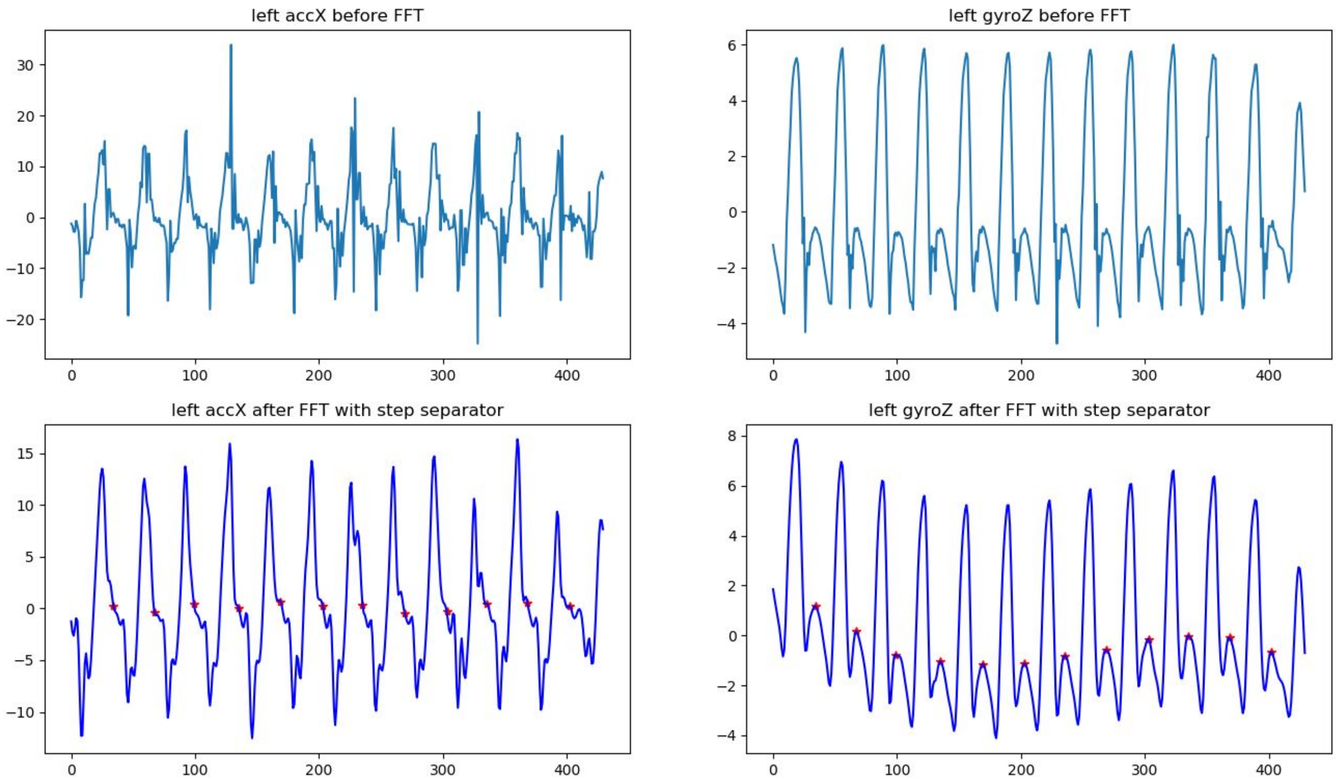
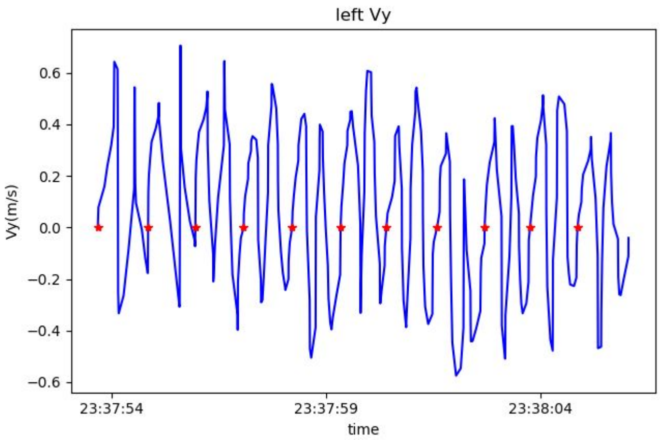


Figure of comparison before and after FFT



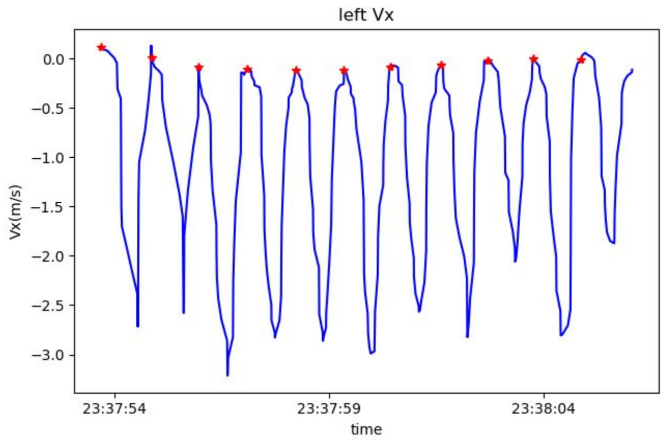


Figure of Vy and Vx

################

left frequency: 0.906976744186

right frequency: 0.882352941176

Test summary:

Testing length of each step...

left normal test p-value: 0.812

right normal test p-value: 0.14

t-test p-value: 0.552

Testing duration of each step...

left normal test p-value: 0.572

right normal test p-value: 0.454

t-test p-value: 0.375

mann whitney u test of step-length of left and right: 0.162

mann whitney u test of step-duration of left and right: 0.136

left walking speed: 1.1

right walking speed: 1.04

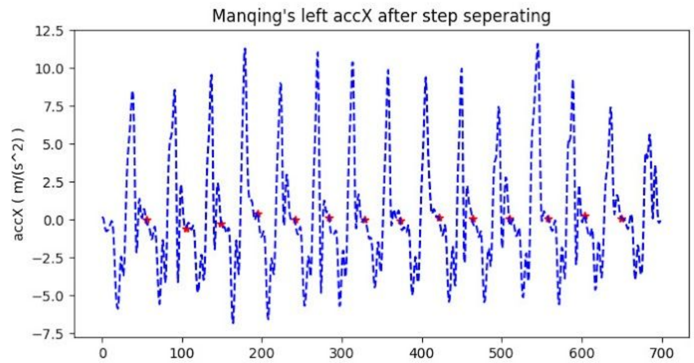
actual walking speed: 1.19

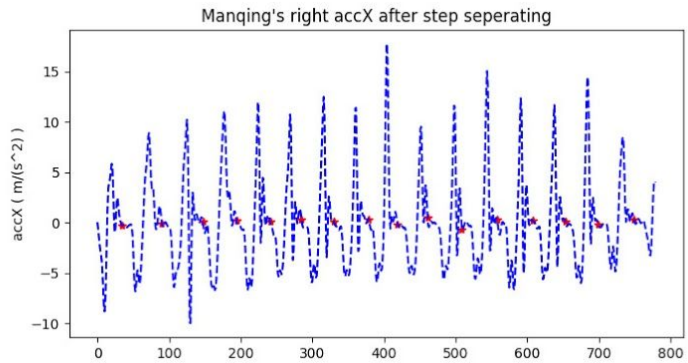
################

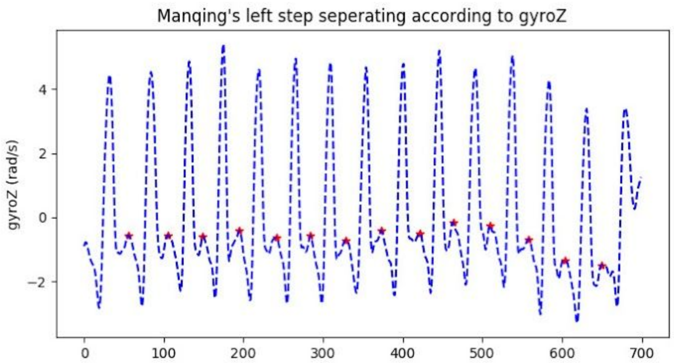
From the results above, we can get the walking pace is about 119 steps/minute and the walking speed is about 1.06 m/s. The walking pace is calculated by FFT and it is pretty reliable. The accuracy of walking speed is 89.1%. From the test summary, we can conclude that we cannot determine that the gait is asymmetric.

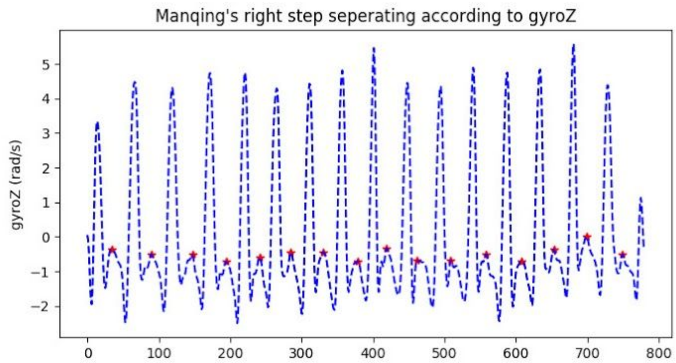
Dataset 2:

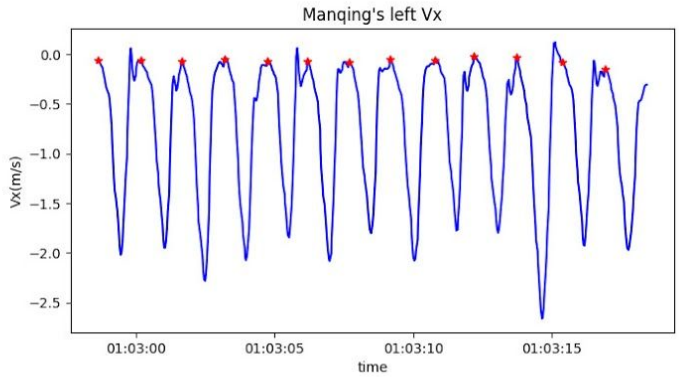
Tester: Manqing Zhu, Female, 20, no injury

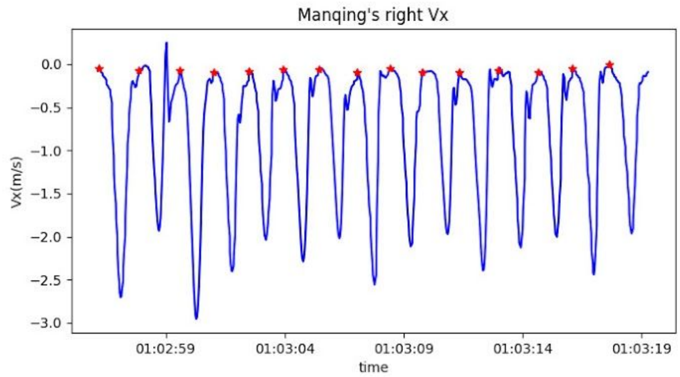












All the data used in the above figures are after noise filtering.

################

left frequency: 0.642857142857

right frequency: 0.653846153846

Test summary:
Testing length of each step...
left normal test p-value: 0.0815
right normal test p-value: 9.773e-06
t-test p-value: 0.707

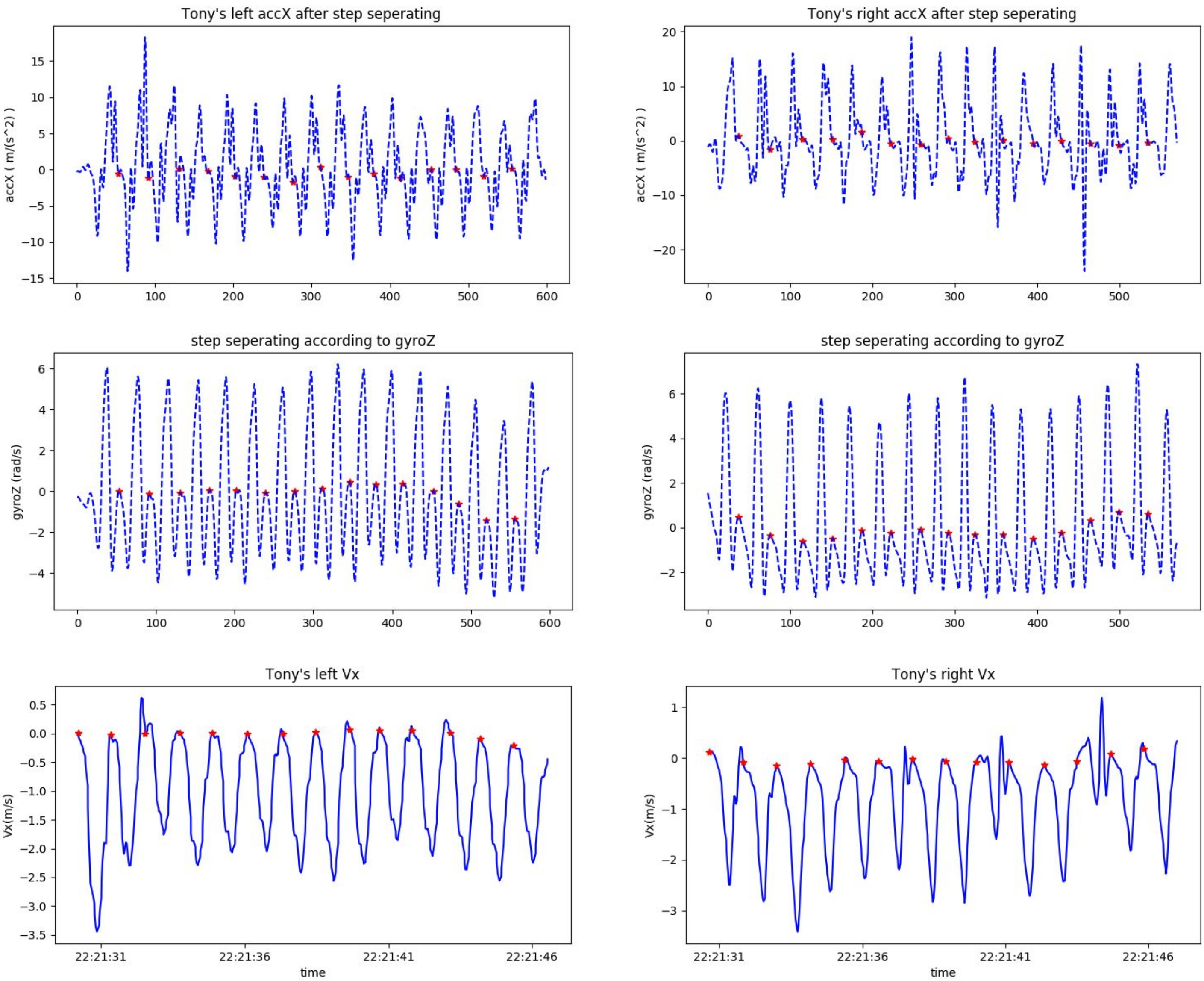
Testing duration of each step...
left normal test p-value: 0.887
right normal test p-value: 0.097
t-test p-value: 0.423

mann whitney u test of step-length of left and right: 0.379
mann whitney u test of step-duration of left and right: 0.234

left walking speed: 0.748
right walking speed: 0.748
actual walking speed: 0.768
################
From the results above, we can get the walking pace is about 77.8 steps/minute and the walking speed is about 0.748 m/s. The walking pace is calculated by FFT and it is pretty reliable. The accuracy of walking speed is 97.32%. The duration of each step for both left and right foot is normally distributed. And it passed the t-test, which indicates that we can confidently conclude that the mean of step durations for left and right foot are equal. The step length of left foot seems to be significantly non-normally distributed. It's understandable considering the fact that the number of sample data isn't big enough ($n \leq 20$) due to the limitation of the testing conditions. Both the step length and the step duration of two feet pass the u-test, which suggests that the shapes of these data distributions of two feet are very similar. Besides, the step frequency of left foot is very close to that of right foot. From the test summary, we can conclude that we cannot determine that the gait is asymmetric.

Dataset 3:

Tester Zhizhou Jiang, Male, 20, Disabled



On the above figures, all the x axis are the index of the data points, the RED dots show the time of the mid-stance. All the plots are after FFT.

################

left frequency: 0.85

right frequency: 0.842105263158

Test summary:

Testing length of each step...
left normal test p-value: 0.000559617599414
right normal test p-value: 0.409655571637
t-test p-value: 0.0788577231529

Testing duration of each step...
left normal test p-value: 0.974264882175
right normal test p-value: 0.591992233442
t-test p-value: 0.616941116352

mann whitney u test of step-length of left and right: 0.0738991919514
mann whitney u test of step-duration of left and right: 0.272735580726

left walking speed: 1.008
right walking speed: 0.805
actual walking speed: 1.0
################
From the results above, we can get the walking pace is about 101.5 steps/minute and the walking speed is about 0.9 m/s. The walking pace is calculated by FFT and it is pretty reliable. The accuracy of walking speed is 90.695%. From the test summary, we can conclude that we cannot determine that the gait is asymmetric. But because this tester is disabled, and it shows a 0.2 difference on left vs right walking speed, which is the biggest difference between other testers. And by checking the accelerometer data on X axis, this tester also shows different data shape between left and right foot.

Conclusion

In this project, we successfully developed an algorithm that is able to calculate the walking pace by Fast Fourier Transform and calculate the step length hence the walking speed by acceleration integration with drift correction. Through statistical tests such as normal test, t test and mann whitney u test, we tried to determine whether the gait is symmetrical or not. Due to limitation of datasets, we were not able to tell how does these attributes vary from different people.

Discussion

1. Problems

There was a problem which we still don’t know how to fix. We did some simplification to the walking model to get a result that is not so strange. We did not consider the rotation of the cell phone when we did the integration to get velocity. So the velocity we got was relative to the coordinates of the phone rather than the vertical and horizontal coordinates. And we assume that in the “mid stance” point the phone is vertical. In consequence the result we got lost some accuracy. We tried to improve our algorithm but could not get a satisfying result.

2. Limitations

Due to limited time and resource, and the fact that the way to clean and extract each dataset is not exactly the same, we only collected 3 set of data which is apparently not sufficient. And due to the reason above we made some simplification to the problem and lost some accuracy in doing so.

We calculated the maximum foot clearance of each step and failed to get a decent result. The reason of the failure might be that unlike displacement in X direction, displacement in Y direction is very small and the sensor and our algorithm is not accurate enough to get a good result.

Analyzing the walking behavior is not an easy job, Apple has a whole team to analyze Apple Watch’s data. Because we have limit time and limit the number of people, we try to collect the data we can really work with, and hopefully get a reasonable result.

We try to do a Chi-Square test on our dataset result, but our data result are numerical not categorical. And the we only have three tester date with limit length, so we can not get/conclude a robust result between different people or vary by age, gender and height.

3. Future improvements

Improvements could be make according to the limitations. Had more time and resources and more research into the problem, we could collect more data sets and see how accurate could the result be. We could also polish the algorithm to get a best result.

With more data, we could also dig into more questions and make use of more tools. We could probably apply machine learning methods and predict the information of the person with the walking data, like gender, age, weight, height, so on. Attributes could be the step pace, walking speed, maximum foot height and the difference between two legs and so on.

We could probably put more thoughts into other questions. For example, is it possible to tell whether someone is not comfortable by his/her gait? Is it possible to tell someone’s personality? Habits? That seems a little impossible, how about the relation between weight and walking pace? Or the relation between height and step length? And maybe some statistic tests could be in place.

Reference

1. Maria Yousefian, “Design and Implementation of a Smartphone Application for Estimating Foot Clearance during Walking”, 2017.
2. A. Laudanski, S. Yang, and Q. Li, “A concurrent comparison of inertia sensor-based walking speed estimation methods,” Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2011, pp. 3484–3487, 2011.
3. Apple Developer Documentation, https://developer.apple.com/documentation
4. An example of butterworth filter, https://stackoverflow.com/questions/12093594/how-to-implement-band-pass-butterworth-filter-with-scipy-signal-butter
5. An example of getting frequency using Fourier Transformation,https://stackoverflow.com/questions/3694918/how-to-extract-frequency-associated-with-fft-values-in-python

Project experience summary

Zhaoyang Li:
Implemented a program that calculates and analyzes walking attributes from gyroscope and accelerometer sensor data collected by cell phone, exercised and improved data collecting, cleaning and analyzing skills, enhanced understanding and ability of utilizing data science methods.

Zhizhou Jiang:
- Experienced collecting Accelerometer and Gyro data from the wearable sensor, converted to a comma separated values file which provided a clean and elegant dataset for the extract-transform-load task
- Participated in data analyze design and data result discussion, gained a deeper understanding of data science research methods and models
- Provided an ideal experimental environment and experimental equipment to ensure the accuracy of experimental results and team development efficiency
- Worked with a team with a version control system (git) and Agile development. successfully solved team conflict, and provided advice for other team members

Manqing Zhu:
- Participated in data collecting and gained a deeper understanding on the gyroscope and accelerometer data from the phone sensor.
- Developed a program of data processing and visualization, implemented the algorithm in paper, experiencing the process of data cleaning and data analyzing.
- Worked as a team with two other classmate, collaborated heavily relying on the Git.