Project report

Title: walking pattern symmetry analysis

Member: Junheng Zhang

Student ID: 301306955

Course: CMPT 353

Instructor: Greg Baker

Outline

Problem thesis:

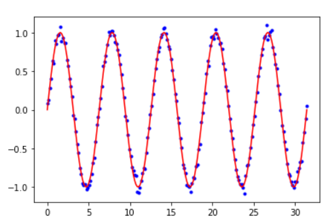
Walking is one of the fundamental movement of humans. It consists movement of feet in three planes, (X, Y, Z). Through the analysis of movement of walking, it allows us to detect dynamic posture problem, body symmetry and injury. In this project, I will be attempting to **analyze walking movement based on accelerations and determine symmetry of user's gait**.

Problem breakdown:

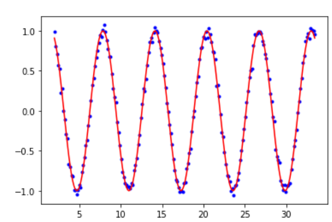
The straight walking movement of symmetry human consist of the following steps:

1. start with standing position with no acceleration on both feet
2. start walking with increasing velocity and acceleration of one foot to lift up in vertical axis (Y axis) and move forward in horizontal axis (X axis)
3. slow down with decreasing velocity and negative acceleration to reach approximately 0 velocity of the foot at the max height of foot lifting in vertical axis (Y axis). In the horizontal axis (X axis), the velocity is also decreased a little with negative acceleration.
4. after reach the maximum height, speed up with increasing opposite velocity and positive acceleration of the foot to land down and move forward in both vertical axis (Y axis) and horizontal axis (X axis).
5. Slow down with decreasing velocity and negative acceleration of the foot in both horizontal axis (X axis) and vertical axis (Y axis) to finish the step and place down the foot.
6. The other axis that is perpendicular to X, Y plane is represented as the Z axis (move away from body or move close to body), which changes similar to the Y axis, but in straight walking the changes would be minor.
7. Repeat the steps above for another foot to continue the walking movement.

As we can see in the steps above, the two foot should be repeatedly moving with the very similar velocity and acceleration in order to walk symmetrically. So for symmetry gait, the change of acceleration of both feet on each dimension should be represented as following:



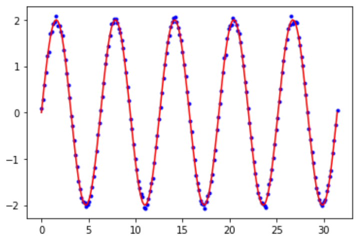
Acceleration for left foot



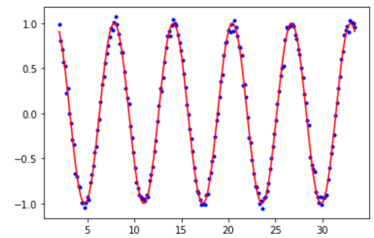
Acceleration for right foot

However, in theory, the changes in acceleration of one foot for each step should show similar pattern to the other foot in terms of symmetry walking, since every step is consisted of accelerations of foot in three dimensions. If there is differentiation between the changes in acceleration on two feet, we would consider it as asymmetric gait (person's two feet are moving differently).

For instance, a person's left foot always steps longer than the right foot when he/she walks. In order to step longer while keeping body balance, his/her left foot should have more acceleration and velocity to keep the body's balance and reach longer displacement in each step. So the acceleration data should look like this:

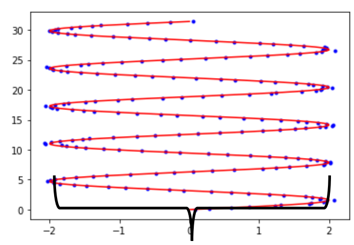


Acceleration for left foot



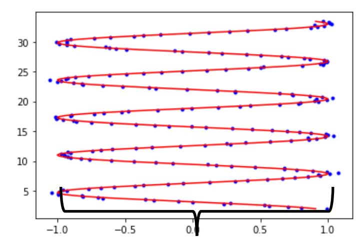
Acceleration for right foot

And for the person with foot or leg injury, he/she would have one foot moving slower than the other, since he/she cannot fully control the injured leg/foot. So he/she would have a narrow range of change in acceleration and velocity for the injured side than the other side. This situation the data will show similar pattern as shown above. However, the acceleration data we obtained always goes up and down around 0 because we have positive acceleration to speed up and negative acceleration to slow down, so the average of the acceleration data should approximately equal to 0. That means we cannot just calculate and compare the population mean (average) of the acceleration. But we are able to differentiate the difference in the spread and distribution of changes in acceleration (the distance from the data to its' mean is different). For example, if we rotate the asymmetrical gait graph above we would see:



Variance

Acceleration for left foot



Variance

Acceleration for right foot

We can easily identify the left foot data having a wider spread of data (from 2 to -2) than the right data (from 1 to -1), thus we can determine the person's gait consists of different changes in acceleration. To compare the spread of acceleration data, the statistic method of variance is utilized to show the distribution of acceleration from mean and the spread of acceleration. This way we will know if the person's foot has more change in acceleration (far from mean -> larger variance) or has less change in acceleration (close to mean -> smaller variance) for asymmetry gait. Although not absolutely identical, similar changes in acceleration can be considered as symmetry gait. Thus, my hypothesis is that **if one person has obvious difference in variance of acceleration, he/she would have asymmetry gait or leg injury.** To prove the validity of my hypothesis, I collected several sets of relatively symmetric walking data to analyze. The test subjects are my parents, my sister, my friend, myself (5 persons for symmetry data) as well as another friend (asymmetry data) who was injured while playing basketball two days ago.

Data collection:

To analyze the variance of acceleration, I gathered walking data from 6 people (5 normal, 1 injured). Every one of us produced data with following steps:

1. Find a field or a path that is straight without slope and obstacle (we used my high school's track field)
2. Used two sports bands for phones to attach the phone and bind around both ankles. Both phones' screen face out and keeps phone portrait and perpendicular to the ground (as the picture shown)

3. Install "Physics Toolbox Sensor Suite" app select "Linear Acceleration"
4. Tap on the plus button to start the record for both side and stand up
5. Start the walk with normal walk speed and try to keep constant speed for both foot while walking. (start the walk within 8 seconds after start record)
6. Walk for around 60 seconds
7. Stop walking and tap the plus button to pause record (stop the record within 8 seconds after stop walking)
8. Repeat the steps above one more time

After the data collection we have 10 normal data sets:

- From my father: R_sensor_normal_1.csv, L_sensor_normal_1.csv, R_sensor_normal_2.csv, L_sensor_normal_2.csv
- From my friend (not injured): R_sensor_normal_3.csv, L_sensor_normal_3.csv, R_sensor_normal_4.csv, L_sensor_normal_4.csv
- From myself: R_sensor_normal_5.csv, L_sensor_normal_5.csv, R_sensor_normal_6.csv, L_sensor_normal_6.csv
- From my mother: R_sensor_normal_7.csv, L_sensor_normal_7.csv, R_sensor_normal_8.csv, L_sensor_normal_8.csv
- From my sister: R_sensor_normal_9.csv, L_sensor_normal_9.csv, R_sensor_normal_10.csv, L_sensor_normal_10.csv

And the data from my injured friend Sam, I was willing to get more data in order to further analyze the injury data set. However, after 3 rounds Sam felt some pain around his left ankle. Sam was no longer able to continue for the experiment; therefore we discontinued his part immediately and let him have more rest. Thus we only received 3 sets of data from Sam for the injury data set. Special thanks to Sam for participating in this experiment.

- From Sam: R_sensor_injury_1.csv, L_sensor_injury_1.csv, R_sensor_injury_2.csv, L_sensor_injury_2.csv, R_sensor_injury_3.csv, L_sensor_injury_3.csv

I also attempted to imitate some asymmetric walking movement with two feet moving slightly different and collected the data as following:

- Left foot has long step: R_sensor_injury_4.csv, L_sensor_injury_4.csv
- right foot has long step: R_sensor_injury_5.csv, L_sensor_injury_5.csv
- Left foot has higher lift: R_sensor_injury_6.csv, L_sensor_injury_6.csv,
- Right foot has higher lift: R_sensor_injury_7.csv, L_sensor_injury_7.csv
- Left foot move outer: R_sensor_injury_8.csv, L_sensor_injury_8.csv
- Right foot move outer: R_sensor_injury_9.csv, L_sensor_injury_9.csv

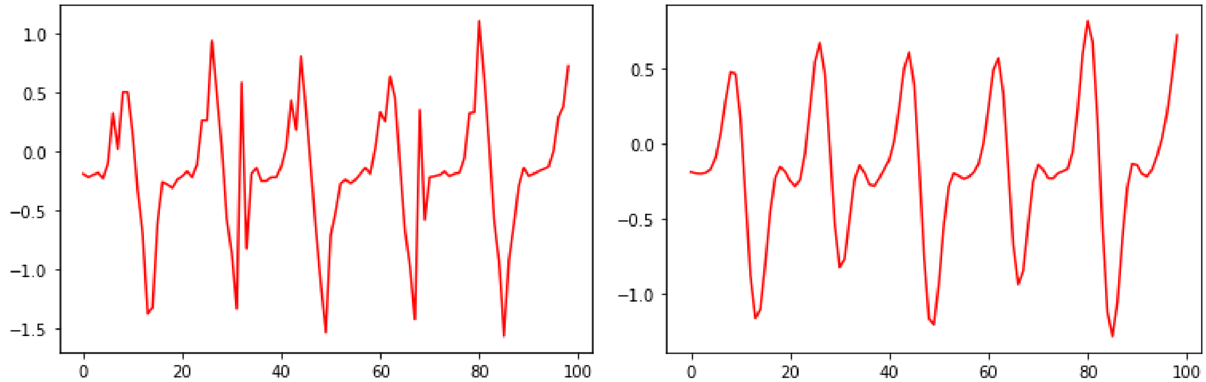
In total, I have 10 normal data sets and 9 injury (asymmetry) data sets to analyze.

Data cleaning

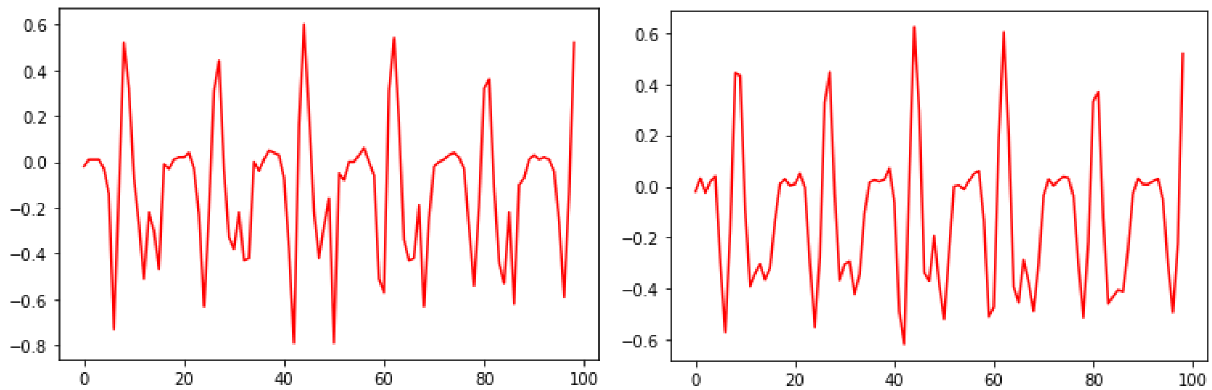
For data cleaning, I used the Butterworth filter as suggested. Because the walking data are all periodic signal and the rhythm for each step is around 0.5s to 1s per step, so I need to determine a proper frequency that suits most walking data. Furthermore, the frequency for the x axis, y axis and z axis are different since the cycle or each axis is different. In theoretical calculation, the sampling rate of the linear accelerator is around 10 sample per second, and I have steps around from 2 steps per second to 1 step per second. Which means the frequency (sample/cycle) should be > 0.1 for X axis. For the Y axis and Z axis, the acceleration has more changes since it always turn to move to opposite sides when the foot reach the max height of each step, so the cycle may be half of X axis and the frequency may be doubled. In order to get the real frequency, I plotted graphs using all the data in 3 axis and compared the original graph and filtered data. One thing I noticed is that for all the data, the acceleration of each axis are distributed in relatively small amount of range, especially for Y axis and Z axis. So when I tested the filter, I was trying to use larger frequencies to keep more original data rather than having less noise because the small frequencies may filter out more truth. Finally, I got the 0.35 as the frequency for the X axis, and 0.75 for the Y axis. However, for the Z axis the values are unstable and in small range, it is hard to filter out the noise from truth. So I estimated the frequency to be 0.8 for the Z axis to keep more original data.

Here are the partial data before and after filtered:

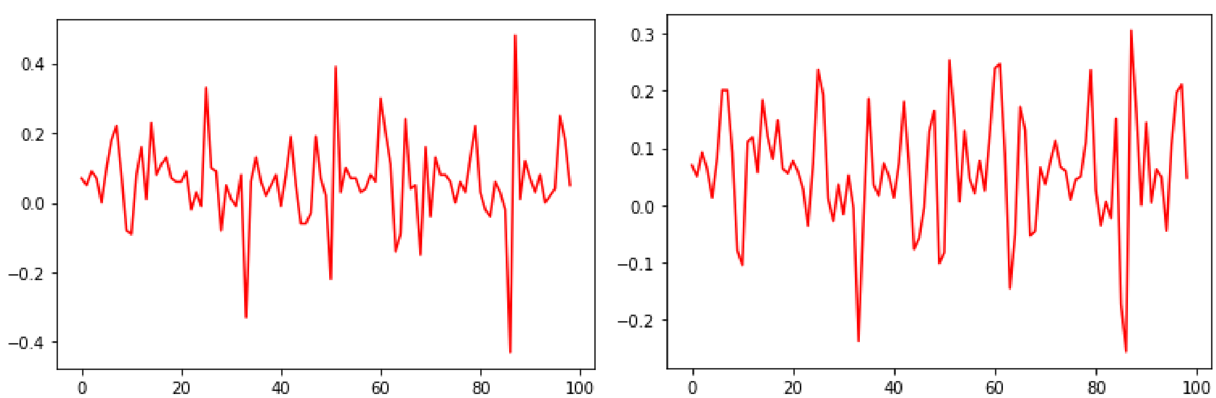
- Original X axis and filtered X axis



- Original Y axis and filtered Y axis:

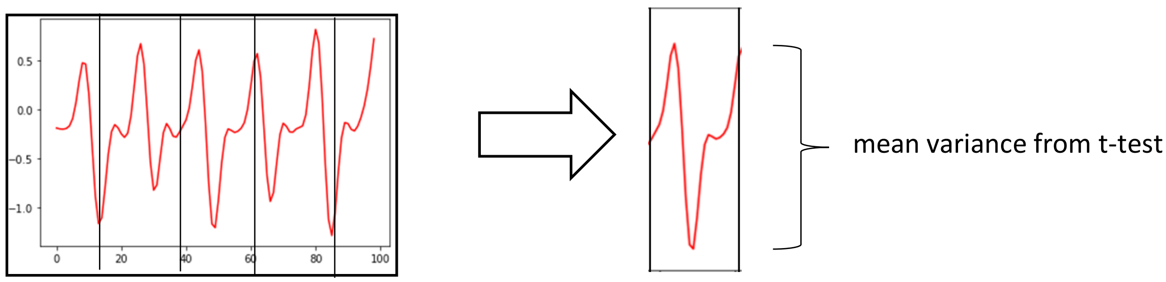


- Original Z axis and filtered Z axis:



Comparing data

To compare the variance, I need to use statistic test. Initially I was planning to use the two-sample t-test to check the population mean of variance in order to see if there the variance from two legs can be differentiated for normal people and injured people. The significance of doing the t-test for the variance sets is to test if the two legs have similar average variance of acceleration so we would know the difference of the distribution of acceleration for two feet. However, the two-sample t-test needs more sample data and two conditions. So I first divide the whole walk data for each data set to many sub groups with 2 seconds data (20 data for each group), so we have at least 30 (60 seconds / 2 seconds) groups of samples for each walk data. The reason to divide into two second intervals is because we walked at 0.5 to 1 second per step, 2 second of walk data will contains at least 1 to 2 complete steps, so we can calculate the variance for each 1 to 2 steps. And the population mean from the t-test can represent the average spread of acceleration for every step.



After that we need to check for the condition for t-test. The first condition is two samples should be independent to each other, which we can know there is no relation between the two samples data because we collected data from different foot independently. The second condition is the two samples data for the t-test have to be normally distributed. In theory, if we have the straight walking data without much changes in the overall walking speed, we should have very similar variance of acceleration which is slightly greater or less than the mean variance of acceleration because the movement and change in acceleration are constant and similar for each steps. I also calculated the normality for each axis from both legs, the results are the following:

X axis	Y axis	Z axis
0.7940935621377686	0.7296659352285054	0.3985011263653246
0.42966182677074205	0.8775784543260584	0.03823540850085658
0.9148615629397007	0.9084894564271877	0.15184062444905239
0.0006083252008000242	0.7460587972533814	0.38791381945545633
0.3899234065870893	0.11636713567686892	0.05773393535031236
0.11909467929134214	0.6684964546884553	0.00016889419378186927
0.6556347857234561	0.7404920968391655	0.10820821228829781
0.24136412906529772	0.037176917967871384	3.389621663816919e-08
0.6789212406571348	0.23588594951383904	0.9220295762174882
0.9255791172880592	0.6924787754952613	0.1473918738691969
0.020957033764581225	0.3767512174918737	0.10261015108106103
0.49678978069590995	0.831060455299923	0.3101272804239186
0.3589546579594381	0.911985577102812	1.6786591807682397e-06
0.47422928394483055	0.3438357667337181	0.4736748075443291
0.08194755111046906	0.013197573504588363	0.07485822045677837
0.7264752639038004	0.0001743821087973174	0.0004235476435777315
0.31703344089729524	0.007523142431881098	1.1127223350541812e-08
0.09664583010594852	0.1697724062852262	0.009217150987812183
0.22030426396904257	0.5975532409613451	7.862552980339487e-06
0.588725012204999	0.2327742198840323	0.002933029469011527
0.2295812661490548	0.00979680069737944	1.4099508532467496e-05
0.87793593861806	0.0009037864976364464	2.8045337358659843e-07
0.010358080799589388	0.08911188210556766	0.17960933764198123
0.04216259592079898	0.016226724987354394	0.0015864160704511733
0.21212879411359511	0.000479258051245831	8.612774198305781e-07
0.009454677189860825	0.032293905783124834	0.08869912019539591
0.2809247568761467	7.30989369774468e-05	0.024521383869390706
0.376780563102964	0.010318455946924128	0.005275264952739423
0.2517375125444765	0.034264926641086706	0.39525676966375817
0.7950451659099914	0.06617258545555672	1.3832461684114596e-05
0.34360803784031446	0.0004342922282384674	0.31327136990000937
0.5896453933210669	0.48495473889716323	4.240741001715153e-05
0.026268682031198434	0.7966422478499601	0.03169441447042107
0.017683806565714926	0.7397959791796669	0.9653498622309589
0.39974217500204445	0.09868033987603725	0.19983444035609732
0.054771548901424945	7.436328855122163e-05	0.40453162178790203
0.1485656484578471	0.010782358652627316	0.13863649455080884
0.9298270901805264	0.08545950603347839	5.759124071604088e-07

assume alpha to reject the normal distribution = 0.05, there are 31/ 38 (81.57%) in x axis, 25/38 (65.79%) in y axis and 20/38 (52.63%) in z axis variances are normal. For those data sets with p-values smaller than 0.05, the possible reasons could be inconsistent speed or too small variance. However, as the most variance data sets are normal, we can try to use the two-sample t-test for the variances. The results from the t-test are the following:

normal_X	normal_Y	normal_Z	injury_X	injury_Y	injury_Z
0.598657	0.946375	0.000811	4.38E-12	0.300684	0.067906
0.466608	0.371771	0.00094	4.15E-07	0.493245	0.006879
0.93226	0.041007	0.003139	2.62E-15	0.004864	0.001527
0.992177	0.386671	0.058497	5.22E-09	3.11E-05	6.16E-05
0.712076	0.796697	0.060769	7.69E-09	0.050378	0.003922
0.691842	0.977938	2.77E-06	0.701006	1.29E-08	1.56E-08
0.841867	0.756464	0.031103	7.83E-05	3.19E-18	6.08E-13
0.279225	0.231814	0.106759	0.000367	8.55E-07	5.81E-08
0.951463	0.001584	0.798846	1.51E-07	0.469207	5.91E-11
0.3955	0.686172	0.490126			

To be clear, the null hypothesis for the t-test is $\mu1=\mu2$ (where $\mu1$ is the population mean of variance of the left foot and $\mu2$ is population mean of variance of the right foot), and the alternative hypothesis is $\mu1\neq\mu2$, and we assume the α value for rejecting null hypothesis is 0.05. The result shows that:

- For normal walking data:
    - X axis: 10/10 (100%) p-values are > 0.05, all fail to reject null
    - Y axis: 8/10 (80%) p-values are > 0.05, most cases fail to reject null
    - Z axis: 5/10 (50%) p-values are > 0.05, half cases fail to reject null
- For injury data from Sam:
    - X axis: 0/3 (0%) p-values are > 0.05, all reject null
    - Y axis: 2/3 (66.7%) p-values are > 0.05, most cases fail to reject null
    - Z axis: 1/3 (33.3%) p-values are > 0.05, most cases reject null
- For asymmetry of longer or shorter:
    - X axis: 0/2 (0%) p-values are > 0.05, all reject null
    - Y axis: 1/2 (50%) p-values are > 0.05, half cases fail to reject null
    - Z axis: 0/2 (0%) p-values are > 0.05, all reject null
- For asymmetry of lift higher or lower:
    - X axis: 1/2 (50%) p-values are > 0.05, half cases reject null
    - Y axis: 0/2 (0%) p-values are > 0.05, all reject null
    - Z axis: 0/2 (0%) p-values are > 0.05, all reject null
- For asymmetry of lift outer or closer:
    - X axis: 0/2 (0%) p-values are > 0.05, all reject null
    - Y axis: 1/2 (50%) p-values are > 0.05, half cases fail to reject null
    - Z axis: 0/2 (0%) p-values are > 0.05, all reject null

From the result, it is apparent that for symmetrical data, we have failed to reject null in most cases, except for the z axis. On the other hand, the injury data and their corresponding asymmetric axis (x for longer or shorter, y for lift higher or lower and z for outer or closer) all have the p-values that are much smaller than the significance level (α value), where shows the obvious differences for the variance of acceleration in asymmetrical data. Because the measurement error of this experiment is relatively large, we might further adjust the significance level (α value) to get more allowance for significant data in symmetry data while keeping rejecting the asymmetric data. However, the current data we have is not sufficient to determine proper α value and establish the confidence interval (how confidence we are for the result). In order to get a conservative and useful result, I compared all the p-value and made the estimated guess of the possible value of α. Now we have $\alpha = 0.05$ for x axis, 0.01 for y axis and 0.001 for the z axis (as z axis has more noise and unstable, will be discussed in limitation). So we get the result as following:

- For normal walking data:
  - X axis: 10/10 (100%) p-values are > 0.05, all fail to reject null
  - Y axis: 9/10 (90%) p-values are > 0.01, most cases fail to reject null
  - Z axis: 7/10 (70%) p-values are > 0.001, most cases fail to reject null
- For injury data from Sam:
  - X axis: 0/3 (0%) p-values are > 0.05, all reject null
  - Y axis: 2/3 (66.7%) p-values are > 0.01, most cases fail to reject null
  - Z axis: 3/3 (100%) p-values are > 0.001, all fail to reject null
- For asymmetry of longer or shorter:
  - X axis: 0/2 (0%) p-values are > 0.05, all reject null
  - Y axis: 1/2 (50%) p-values are > 0.01, half cases fail to reject null
  - Z axis: 1/2 (50%) p-values are > 0.001, half cases reject null
- For asymmetry of lift higher or lower:
  - X axis: 1/2 (50%) p-values are > 0.05, half cases reject null
  - Y axis: 0/2 (0%) p-values are > 0.01, all reject null
  - Z axis: 0/2 (0%) p-values are > 0.001, all reject null
- For asymmetry of lift outer or closer:
  - X axis: 0/2 (0%) p-values are > 0.05, all reject null
  - Y axis: 1/2 (50%) p-values are > 0.01, half cases fail to reject null
  - Z axis: 0/2 (0%) p-values are > 0.001, all reject null

As seen from the results, we failed to reject my hypothesis of **difference in variance of acceleration correlates with asymmetry gait or leg injury.** Thus we can calculate the user's variance of acceleration of feet movement to conclude whether if the user has asymmetry gait or injury.

Limitations

There are many limitations in this experiment I want to address:

- Measurement error: there are several restrictions that could influence the accuracy of data, causing measurement errors in this experiment. First and foremost, the phones which used to collect data may give different results when we use two phones with different specifications or operating system. Therefore, I suggested to use phones with similar specifications to reduce the measurement error in user guide. Furthermore, the tightness and height level of sports band we used to bind the phone may also affect the data since it may get loose or drop lower when moving fast. Thirdly, the angle of phones to the ground may influence Z axis a lot. Because the Z axis in the straight walking movement is changed in a very small range (sometimes -0.2 to 0.2), and the scale of measurement is relatively large (0.01). So every small difference of angle to the ground can cause large difference of acceleration for each feet. That is the reason I give the Z axis much smaller α than the other axis.

- Too few data sets: because we only have 6 people to collect the data from, the result from this data analysis can be very limited by lack of diversity of data. And we cannot know how confident we are for the result; we cannot predict a confidence interval based on these few sets of data. The symmetry data we have seems not quite symmetry, we do not know how symmetrical the collectors really are, and we can still find some asymmetrical patterns in both y and z axis. On the other hand, the injury data are also somewhat limited for us to analyze since we only collected 3 sets or injury data from one person.

Real application

As we have more data set to compare and analyze, we can have more precise methods of collecting accurate data. We should be able to further develop and practice this app in the following application:

- Phone app for public user: help user to check their body balance and symmetry
- Medical recovery: help patients to identify their recovery state after leg or foot injury
- Athlete: help athlete to identify the muscle symmetry and body balance from their movement
- Battlefield: help commander to detect the injury condition of soldier.

Project Experience Summary:

Analyze human's walk gait to determine the symmetry of human

Design experiment and organize volunteers to collect data for experiment properly.

Use statistical skill to analyze data and determine the significance level.