CMP6046B Ubiquitous Computing Coursework Portfolio

James Burrell 10026300

Contents

1 A3 Activity Recognition

1.1 Stage 2 - Data Analysis and Classification Modelling

Data used has been taken from the A3 Reserve Data file on Blackboard. I have used the data collected under the filename 'aks249' with permission from one of the module teaching assistants.

I had collected my own set of data but unfortunately the smartphone I was using did not have a barometer, which I was unaware of whilst I carried out data collection.

Table 1 shows a list of the activities that were carried out and their corresponding index when displayed in plots.

Activity	Index
Stationary	0
Walking-flat-surface	1
Walking-up-stairs	2
Walking-down-stairs	3
Elevator-up	4
Running	5
Elevator-down	6

Table 1: Activities and their indexes.

1.1.1 To Do 1

"Export and save this plot as Fig. 1 Raw Data."

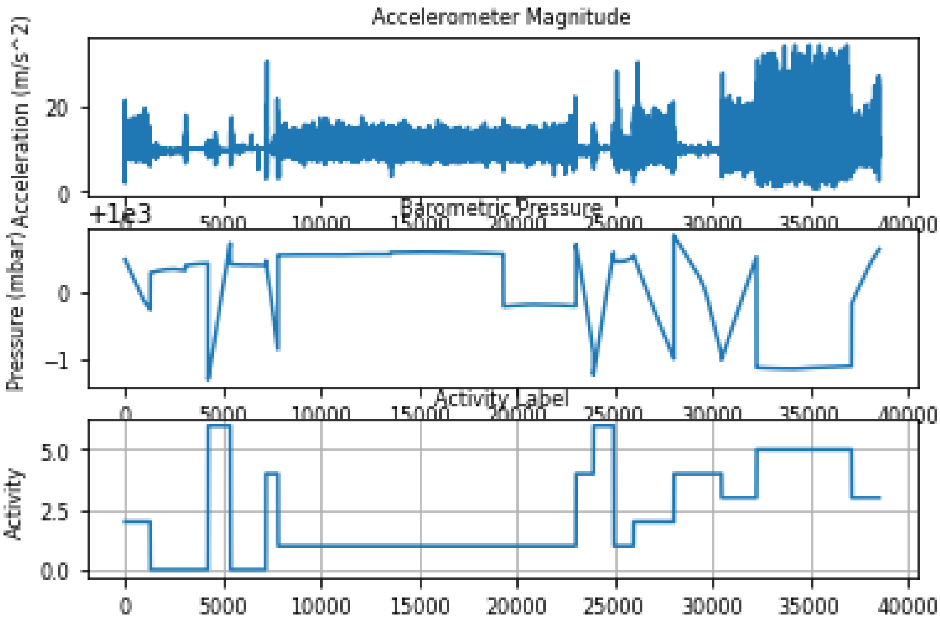

Figure 1: Raw data.

1.1.2 To Do 2

"Export and save this plot as Fig. 2 Feature Data but also change the index in plt.plot(features[:, index]) to see other extracted features. Export these also and don't forget to change the plot title."

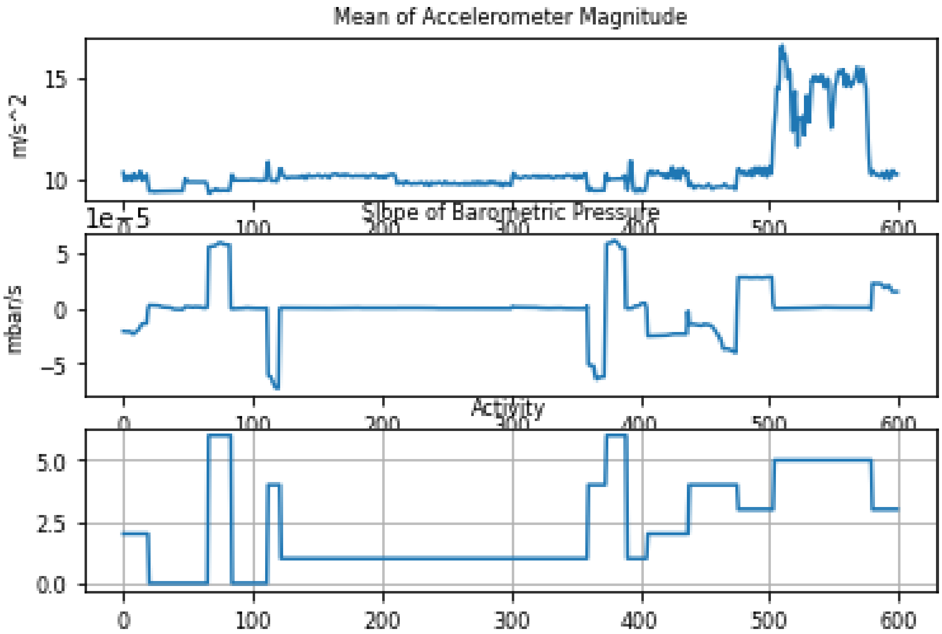

Figure 2: Mean of accelerometer magnitude and slope of barometric pressure.

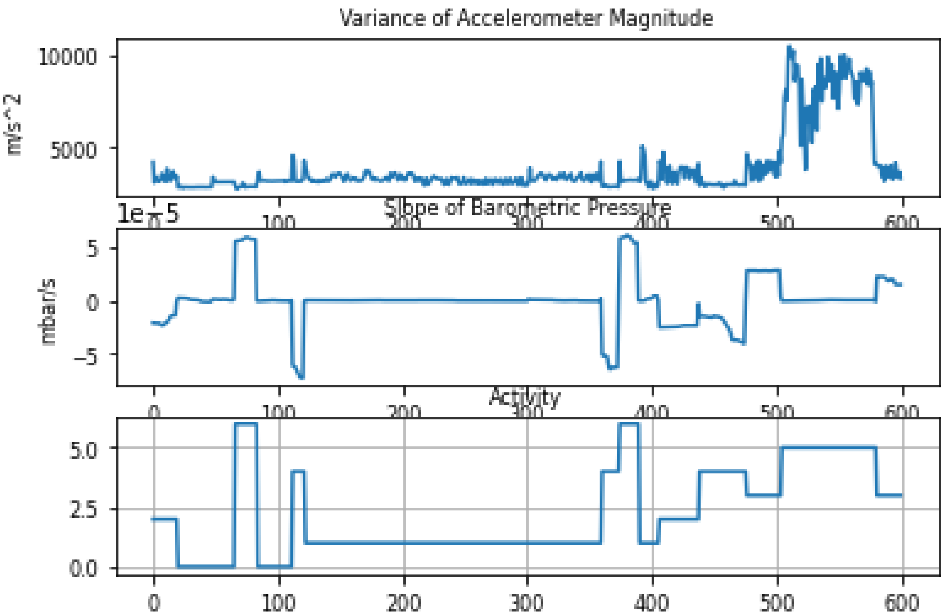

Figure 3: Variance of accelerometer magnitude and slope of barometric pressure.

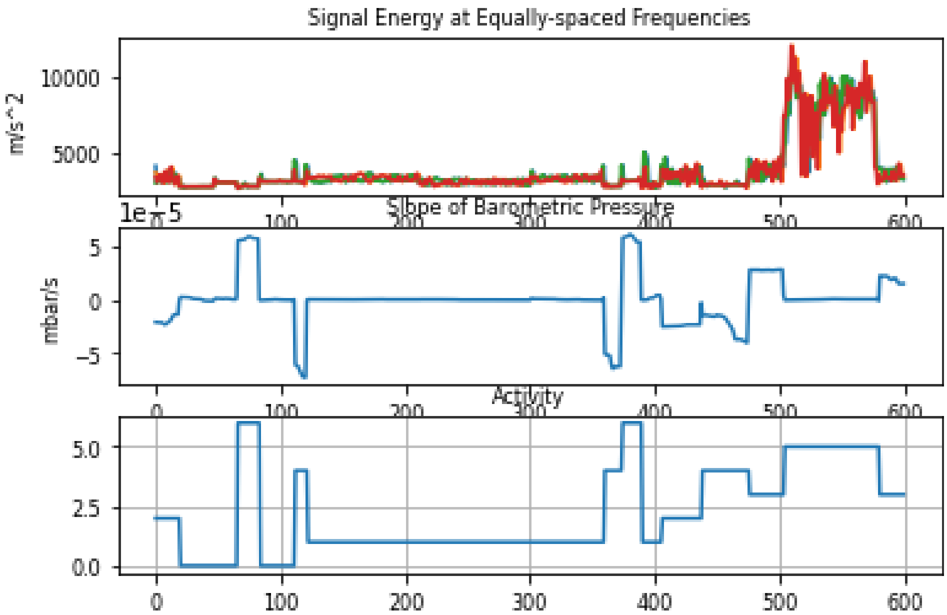

Figure 4: Signal energy at equally-spaced frequencies and slope of barometric pressure.

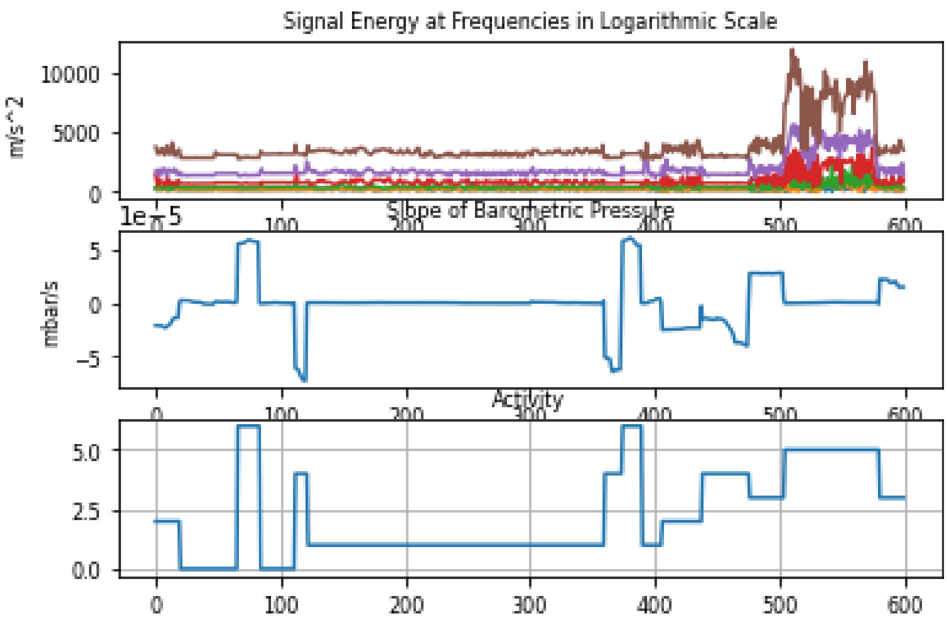

Figure 5: Signal energy at frequencies in logarithmic scale and slope of barometric pressure.

1.1.3 To Do 3

"Comment on how each of the features could inform on which activity is happening. Use your plots as a reference. Don't forget to use the notes you made about your location and phone placement during the data collection stage if this had an effect."

Accelerometer Magnitude

Magnitude of acceleration helps to distinguish between all activities. Large changes in accelerometer magnitude suggest activities with large ranges of movement in both the up and down direction and the forward and back direction, such as running.

This is shown in the first plot in Figure 2, where the further right lines are large, suggesting high changes in accelerometer magnitude. When compared to the third plot, the corresponding activity plot indicates the activity was interpreted as running.

Small changes in accelerometer magnitude suggest activities in which the range of motion in both the up and down direction and forward and backward direction is lower but not zero, such as walking. Very little change in accelerometer magnitude suggests the person is still.

Slope of Barometric Pressure

Firstly, barometric pressure is a feature used to differentiate activities in which changes in altitude are significant, for example, walking up and down stairs.

Decreases in barometric pressure suggest the person is moving downwards, increased suggest they are moving upwards. Relatively no change in barometric pressure or constant fluctuation suggest activities that take place at a single altitude, such as being stationary or walking/running on a flat surface respectively.

Slope of barometric pressure is an extracted feature which better differentiates between activities that involve changes in altitude. It indicates a gradient which equates to the difference in altitude over time.

In the second plot in Figure 2, positive values indicate downward movement (due to air pressure increasing as you move closer to the ground) and negative values indicate upward movement (due to air pressure decreasing as you move away from the ground).

The higher the magnitude of these values, the faster the rate of change in altitude in the up or down direction. This aids in differentiating between all activities but particularly running and elevator up/down, in which both show large changes in barometric pressure, but not how quickly they each changed.

This is confirmed in Figure 2 where we can see that areas where the slope of barometric pressure is highest correspond to activity 6, elevator-down, and where the slope is lowest it corresponds to activity 4, elevator-up. In addition, we can see that activities that involve slower rates of change in altitude produce lower slopes, for running produces a slope of approximately zero.

Variance and Mean of Accelerometer Magnitude

These extracted features help differentiate activities through showing the variation in acceleration and averaging the accelerometer magnitudes. This particularly helps in differentiating between stationary, walking and running activities, as the magnitude of acceleration is displayed more clearly.

In the first plot in Figure 2, stationary activities give a clear average of approximately 9.8m/s^2, solely due to the pull of gravity and no other movement. All forms of walking give higher magnitudes of acceleration on average compared to stationary, but walking up/down stairs give even higher values due to the additional change in acceleration in the up and down direction.

As the data I used was reserve data collected by an unknown user and did not come with any notes, I am unable to comment if the locations or phone placement had an effect on the results.

1.1.4 To Do 4

"Using the values in the confusion matrix to compute the precision, recall, and accuracy scores for each of the activities."

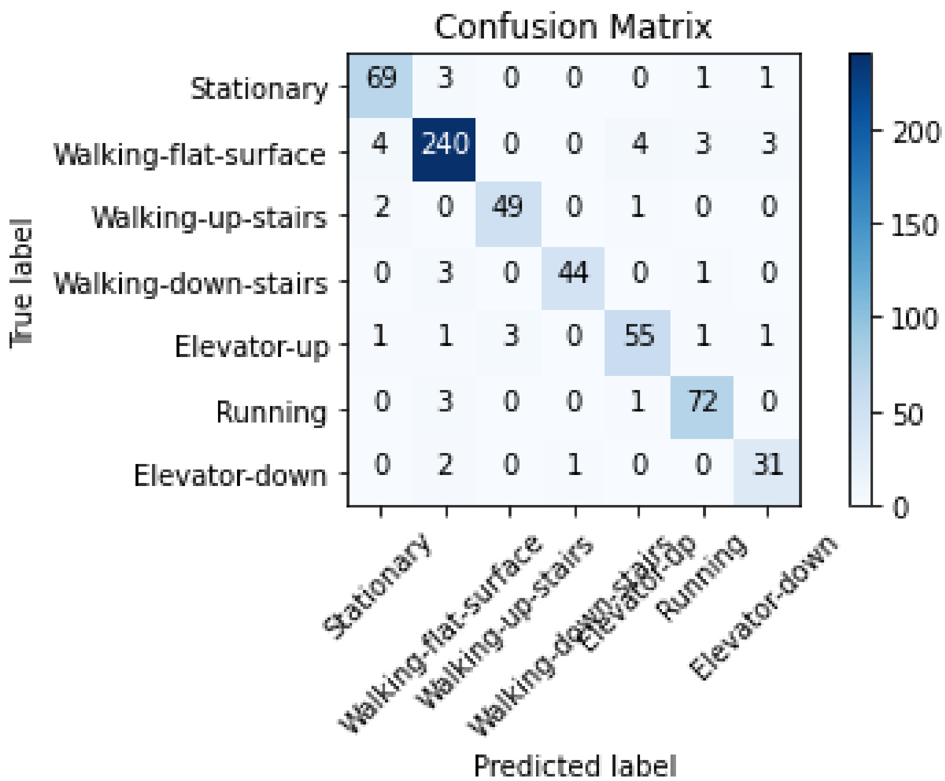

Figure 6: Confusion matrix (single user).

Activity	Precision $P = \frac{T_p}{T_p + F_p}$	Recall $R = \frac{T_p}{T_p + F_N}$	Accuracy $A = \frac{T_p + T_N}{T_p + F_p + T_N + F_N}$
Stationary	$\frac{69}{69+7} = 0.91$	$\frac{69}{69+5} = 0.93$	$\frac{69+491}{69+7+491+5} = 0.98$
Walking Flat Surface	$\frac{240}{240+12} = 0.95$	$\frac{240}{240+14} = 0.94$	$\frac{240+320}{240+12+320+14} = 0.96$
Walking Upstairs	$\frac{49}{49+3} = 0.94$	$\frac{49}{49+3} = 0.94$	$\frac{49+511}{49+3+511+3} = 0.99$
Walking Downstairs	$\frac{44}{44+1} = 0.98$	$\frac{44}{44+4} = 0.92$	$\frac{44+516}{44+1+516+4} = 0.99$
Elevator Up	$\frac{55}{55+6} = 0.90$	$\frac{55}{55+7} = 0.89$	$\frac{55+505}{55+6+505+7} = 0.98$
Running	$\frac{72}{72+6} = 0.92$	$\frac{72}{72+4} = 0.95$	$\frac{72+488}{72+6+488+4} = 0.98$
Elevator Down	$\frac{31}{31+5} = 0.86$	$\frac{31}{31+3} = 0.91$	$\frac{31+529}{31+529+5+3} = 0.99$

Overall Accuracy: 0.933

Figure 7: Confusion matrix metrics (single user).

1.1.5 To Do 5

"Calculate the precision, recall and accuracy scores for all activities. Compare this with the results from the results in 5, comment on the differences."

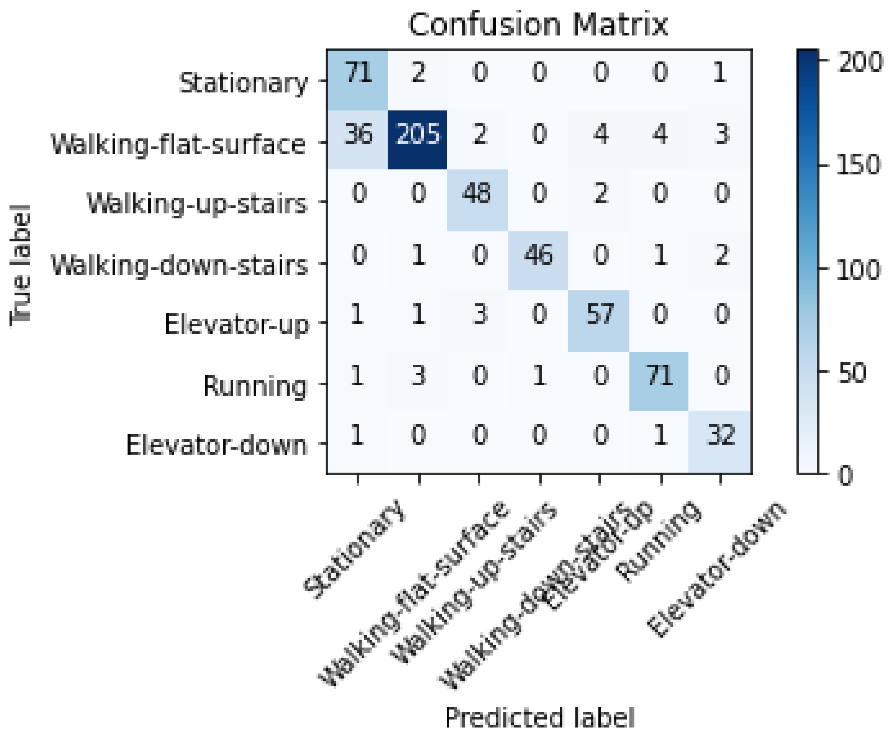

Figure 8: Confusion matrix (all users).

Activity	Precision $P = \frac{T_p}{T_p + F_p}$	Recall $R = \frac{T_p}{T_p + F_N}$	Accuracy $A = \frac{T_p + T_N}{T_p + F_p + T_N + F_N}$
Stationary	$\frac{71}{71+39} = 0.65$	$\frac{71}{71+3} = 0.96$	$\frac{71+459}{71+39+459+3} = 0.93$
Walking Flat Surface	$\frac{205}{205+7} = 0.97$	$\frac{205}{205+49} = 0.81$	$\frac{205+325}{205+7+325+49} = 0.90$
Walking Upstairs	$\frac{48}{48+5} = 0.91$	$\frac{48}{48+2} = 0.96$	$\frac{48+482}{48+5+482+2} = 0.99$
Walking Downstairs	$\frac{46}{46+1} = 0.98$	$\frac{46}{46+3} = 0.92$	$\frac{46+484}{46+1+484+3} = 0.99$
Elevator Up	$\frac{57}{57+6} = 0.90$	$\frac{57}{57+5} = 0.93$	$\frac{57+473}{57+6+473+5} = 0.98$
Running	$\frac{71}{71+6} = 0.92$	$\frac{71}{71+5} = 0.93$	$\frac{71+459}{71+6+459+5} = 0.98$
Elevator Down	$\frac{32}{32+6} = 0.84$	$\frac{32}{32+4} = 0.94$	$\frac{32+498}{32+6+498+4} = 0.98$

Overall accuracy: 0.833

Figure 9: Confusion matrix metrics (all users).

Comparing the calculations in Figure 7 to Figure 9 shows the confusion matrix generated from all users data have varying levels of increase or decrease between metrics of precision, recall and accuracy. For example, Figure 7 shows that the precision, recall and accuracy values for the activity labeled 'Walking Flat Surface' are 0.95, 0.94 and 0.96 respectively. Compare this to Figure 9 with the same activity, the values are 0.97, 0.81, and 0.90.

However, the overall change in metrics across all activities suggest that when data collected by all users is compared, the metrics of precision, recall and accuracy are lower than that of a single user's data. This is supported by the overall accuracy of Figure 7's data being 0.933, whereas Figure 9's data overall accuracy being 0.833.

This goes against the expected outcome of all three metrics increasing with repeated data. One explanation to this could be large differences in data collection methods between users. For example, one user may have collected data for activities whilst holding their smartphone in a hand, whereas another may have placed their smartphone in a pocket.

This could result in different values for accelerometer magnitude and change in barometric pressure, particularly for relatively more vigorous activities such as walking or running.

1.1.6 To Do 6

"Identify the top features and reason whether these features make sense. Make further plots showing these features versus the activities to support your reasoning. Don't forget to use the notes you made about your location and phone placement during the data collection stage if this had an effect."

The model identified the following 5 features as the most important in differentiating between activity types:

1. Mean magnitude of acceleration.
2. Variance of magnitude of acceleration.
3. FFT Power spectrum of equally-spaced frequencies.
4. FFT Power spectrum of frequencies in logarithmic scale.
5. Slope of barometric pressure.

Mean Magnitude of Acceleration

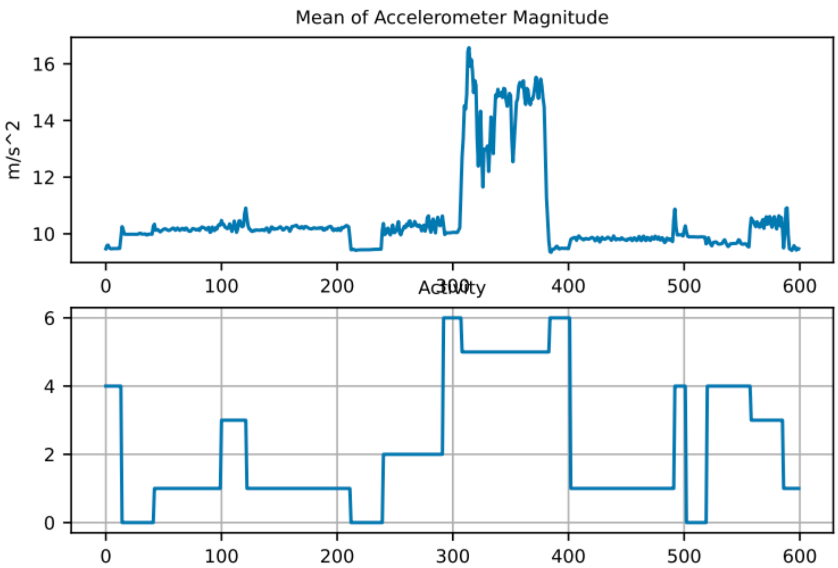

Figure 10: Mean magnitude of acceleration and activity.

The mean magnitude of acceleration is very useful in differentiating between activities as it represents the amount of movement overall in all directions. Figure 10 shows how high values of average accelerometer magnitude correspond to activities with high amounts of movement such as running.

Conversely, activities with lower amounts of movement in all directions, such as walking, have lower values of average accelerometer magnitude, also shown in Figure 10.

Variance of Magnitude of Acceleration

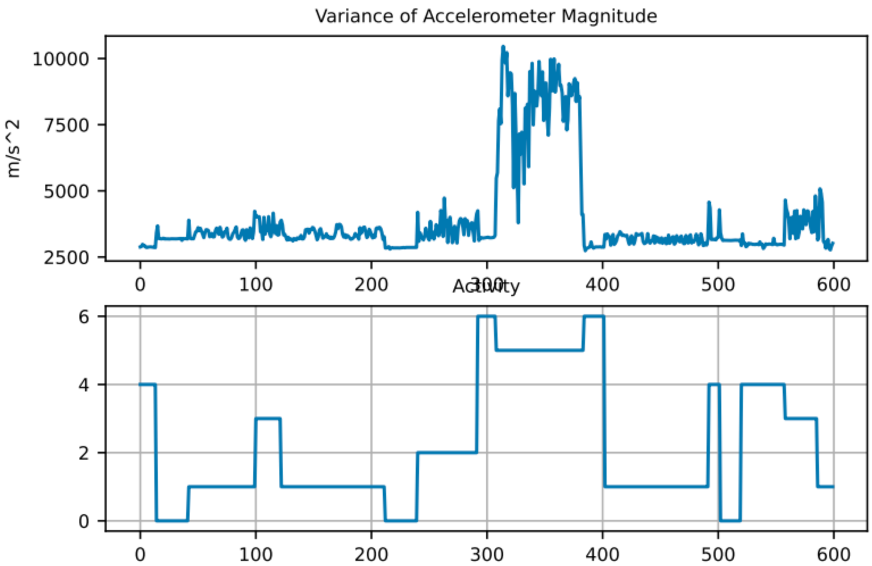

Figure 11: Variance of magnitude of acceleration and activity.

Variance in the magnitude of acceleration helps to differentiate between activities that may have similar values of mean magnitude of acceleration.

For instance, Figure 11 shows both elevator-down and running activities both have high mean acceleration magnitudes compared to other activities. However, Figure shows that the range in variance for running activities is much wider than that of elevator-down.

FFT Power Spectrum (Equal-spaced and Logarithmic Scale Frequencies)

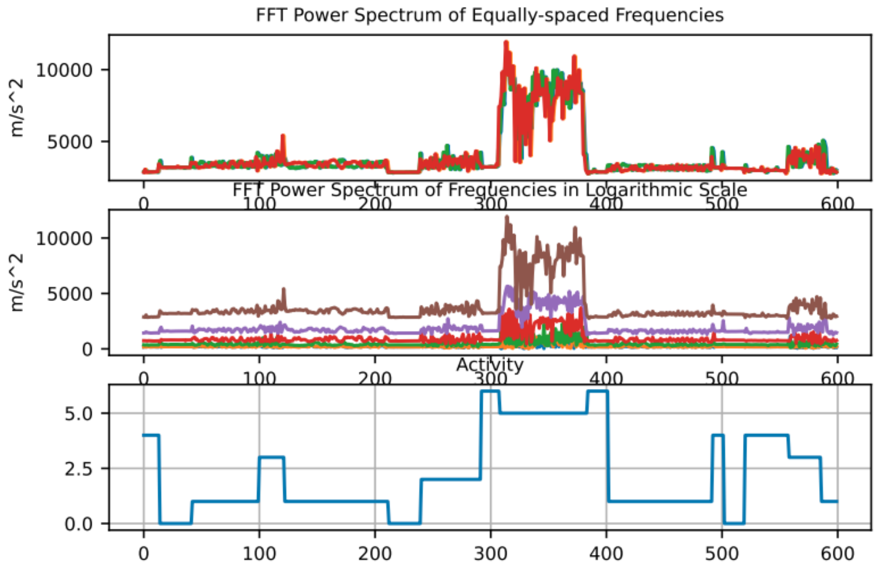

Figure 12: FFT Power spectrum (equal-spaced), FFT Power spectrum (logarithmic scale) and activity.

The power spectrum at equal-spaced and logarithmic scale frequencies use the FFT (Fast Fourier Transform) to remap the accelerometer magnitude data from a time based domain to a frequency domain. This allows for the amplitude of accelerometer values to be viewed against their corresponding frequencies, as seen in Figure 12. Frequencies can then be isolated and interpreted to provide information which aids activity recognition.

Slope of Barometric Pressure

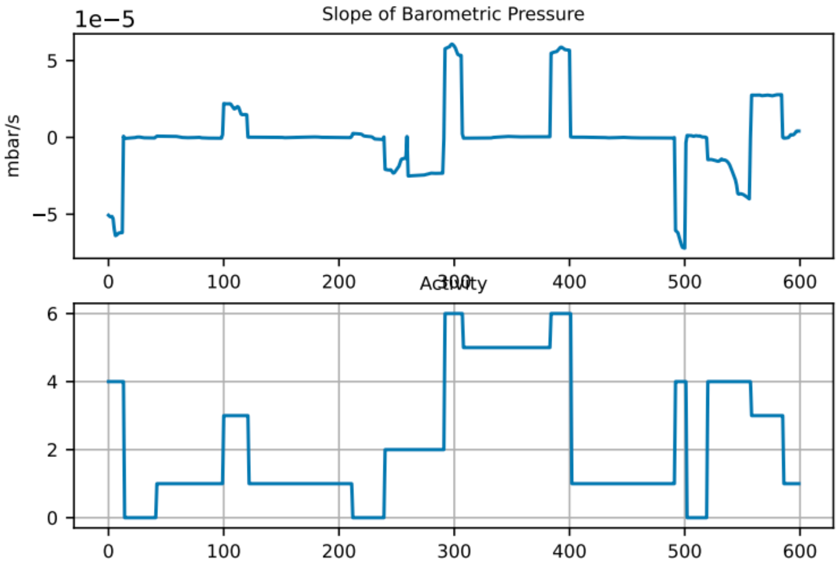

Figure 13: Slope of barometric pressure and activity.

Slope of barometric pressure represents the rate of change in barometric pressure over time. This is useful in differentiating between activities that involve changes in altitude, such as walking upstairs/downstairs, and elevator up/down.

A higher magnitude of slope represents activities with relatively quick increases or decreases in altitude. In addition, the polarity of the slope also can be used to differentiate between upward or downward movement.

Negative slope values correspond to upward movement, and positive slope values correspond to downward movement. This can be seen in Figure 13.

2 A4 Indoor Localization

2.1 Stage 2 - Data Analysis and Classification Modelling

The data that has been used for this assignment has come from the A4 Reserve Data file on Blackboard, with permission from one of the module teaching assistants.

Table 2 shows the seven locations that WiFi hotspot data was collected from. Figure 14 shows the approximate locations of the seven places on the UEA campus.

Number	Code	Description
1	NS102	New Science, room 1.02.
2	NSF01	New Science, first floor foyer.
3	NSFG	New Science, ground floor foyer.
4	JSCG	Julian Study Centre, ground floor foyer.
5	DART	2nd floor SCI, D'Arcy Thompson room.
6	HANE	Outside Hane's office, SCI 2.23.
7	GRAPH	Outside Graphics lab, SCI 2.28.

Table 2: WiFi access point scan locations.

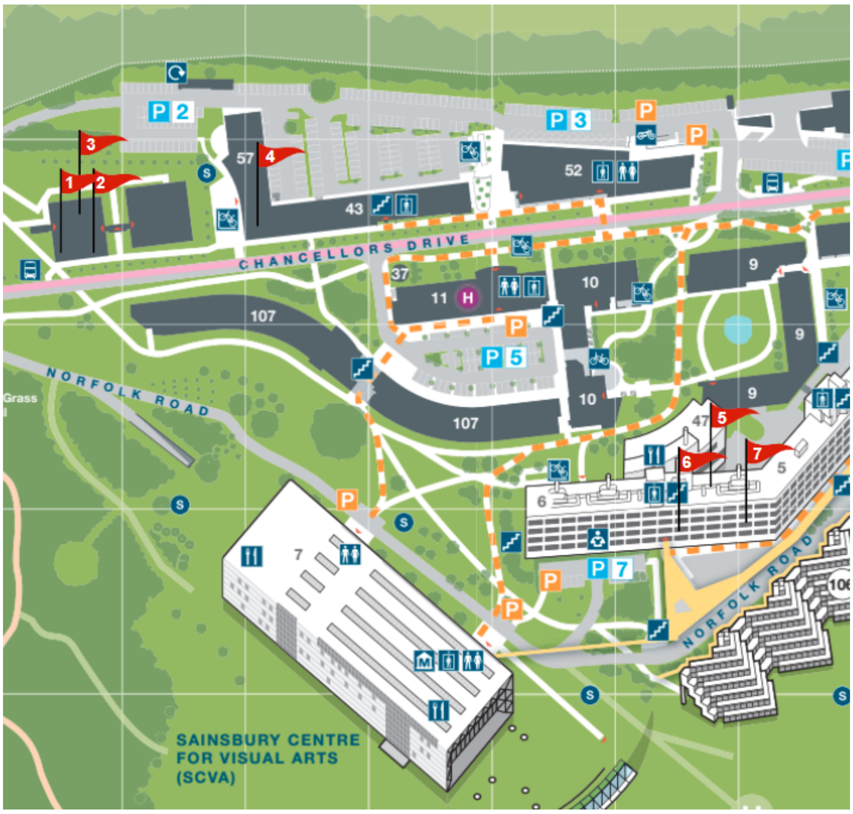

Figure 14: WiFi access point scan locations.

2.1.1 To Do 1

"Export and save these plot as Fig. 1 Signal Data. Comment on what you observe in these plots with regard to the signal strengths and locations."

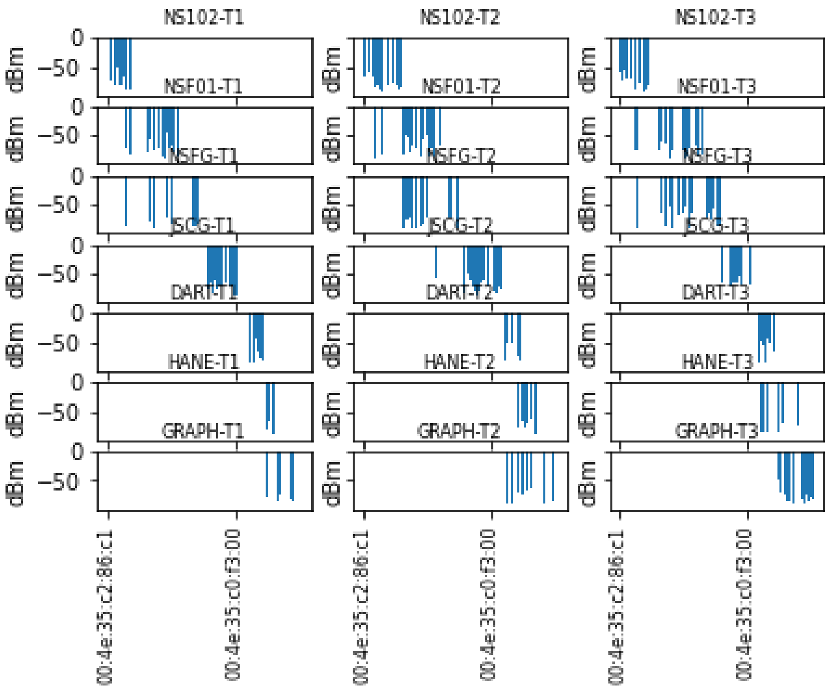

Figure 15: Signal data.

Each subplot in Figure 15 represents an individual WiFi hotspot scan at a given location. Each subplot's title shows the location that the WiFi scan was taken at and is appended with -Tx, with x indicating what trial number the scan was. The x axis lists the different MAC addresses of the WiFi access points that the smartphone connected to during the trials across all locations. The y axis represents the signal strength, measured in decibel-milliwatts (dBm), between the smartphone and a particular WiFi access point.

Each vertical line on the graph shows how strong the signal was between the smartphone and an access point. The more negative the vertical line, the stronger the signal is. MAC addresses that do not have vertical lines indicate that the device is not currently in range of them and therefore have a reading of zero.

From these subplots, we can see which locations share WiFi access points, and get an indication of how close they are together. For example, all three trials scanning at location NS102 show vertical lines in the same x columns (MAC addresses) as all three trials at NSF01, on the far left of the main cluster. At these shared MAC addresses we can see that whilst the signal strengths are similar, there are differences between them in NS102 and NSF01. This could indicate that one location is closer to a specific access point than the other, but are nonetheless in a similar geographical location.

On the other hand, locations that do not share any of the same WiFi access points could be interpreted as being far from each other. For example all three trials at GRAPH show no shared vertical lines (i.e., no shared WiFi access points) with NS102, not even at the lower signal strengths, which could indicate that GRAPH and NS102 are geographically far from each other.

At each of the locations, the three trials have small differences in which WiFi access points were found and also small differences in signal strength at shared WiFi access points. This can be seen in trials T1, T2 and T3 at HANE, for instance. Whilst there are shared access points between all three of the trials, some are missing between trials. One explanation to this could be that the location could represent a large room, with trials taken at opposite ends of the room and/or with the smartphone in a different orientation. This could also be used to explain changes in signal strength at shared WiFi access points.

2.1.2 To Do 2

"In this part of the script write some short code that calculates a Similarity Matrix using the features as input and outputs a square matrix called similarity_matrix. For this you will need calculate the Cosine Similarity between the signal patterns (features) between each of the locations."

For this I have used the cosine_similarity function from the SciKitLearn and passed in the feature arrays as arguments.

2.1.3 To Do 3

"Export and save these plot as Fig. 2 Similarity Data Self. Comment on what you observe in these plots with regard to the similarities and locations"

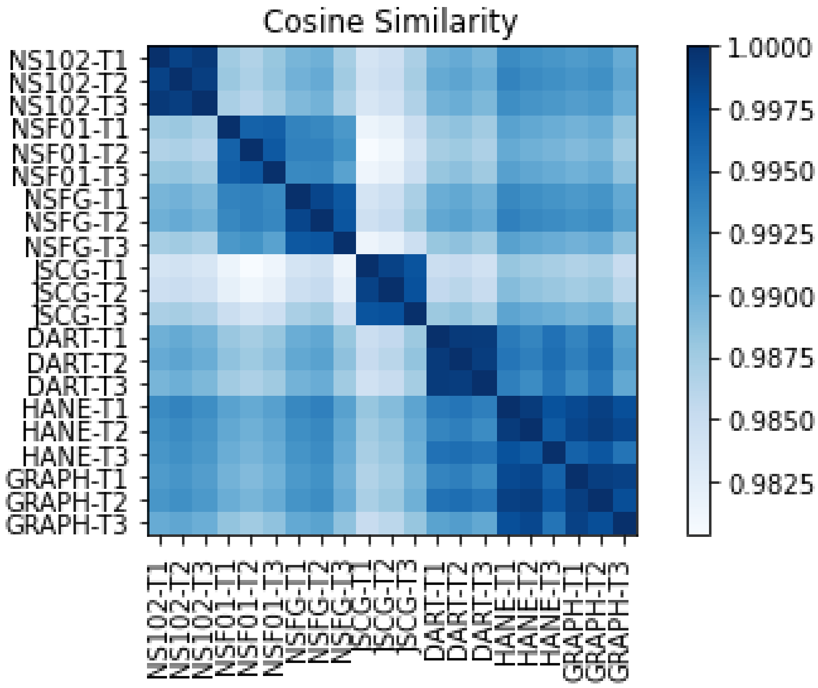

Figure 16: Cosine similarity matrix (single user data).

Figure 16 shows a visualization of a cosine similarity matrix produced using the MatPlotLib library. The cosine similarity of two vectors is a metric used to evaluate the similarity between two sequences of numbers. In this instance, for each trial, the signal strengths from accessed WiFi hotspots at different locations are compared with one another. The closer the signal strength values are to each other, the higher the cosine similarity value, which is then represented as a darker shade of blue on the plot.

The line of darkest blue on the diagonal is where a trial has been compared to itself, and therefore shares all the same access points and signal strengths, resulting in a cosine similarity of 1. These self comparisons do not provide any relevant data other than to prove the cosine similarity calculation function is correct.

During data collection, three scanning trials were carried out for each location. When referring to histograms in Figure 15, slight variances in signal strength and accessed WiFi hotspots can be seen between the trials. This can be seen mirrored in the similarity matrix, where the trials at each location produce a three by three square across the diagonal. The variance in shades of blue in the squares indicates small levels of change in data collected between trials, but are similar enough to show they were most likely taken from the same location.

The cosine similarity matrix (Figure 16) makes visualizing similarities in location easier. For example, the upper left quadrant of the matrix shows an area of on average darker blue, with corresponding cosine similarity values ranging from approximately 0.9875 to 0.9975, excluding self comparisons. This area is where trials from NSF01 and NSFG have been compared. The area of on average darker blue suggests that these locations are geographically close together, and this can be confirmed from the map of trial locations

Likewise, another area on the matrix that suggests two locations are close together is the bottom right quadrant, where trials at HANE and GRAPH are compared. This produces the area of darkest average shade due to their cosine similarity values being high. From this we can similarly determine that HANE and GRAPH are geographically close together, but also that they are closer together than NSF01 and NSFG are, due to the average darker shade.

The area surrounding the cosine similarity comparisons of HANE and GRAPH also show an above average set of cosine similarity values. These are where comparisons between trials at locations HANE and GRAPH have been compared to DART. The lighter shades indicate that DART is geographically similar to HANE and GRAPH, but not as close as HANE and GRAPH are to each other. Again this can be confirmed when considering their actual physical locations, as HANE and GRAPH are rooms on the same floor of the same building, whereas DART is a room on another floor of the same building.

Not all areas of the similarity matrix represent accurate locational data. For instance, the shades of blue in the matrix indicate that the trials conducted at JSCG suggest that it should be geographically further away from all NS locations (NS102, NSF01 and NSFG) than both HANE, GRAPH and DART are from JSCG. As seen on the matrix, when the signal strength values taken at trials at JSCG are compared to locations at NS, they produce the lightest shades of blue on the entire matrix on average. In addition, when HANE, GRAPH and DART are compared to JSCG, they produce areas of darker shades of blue compared to the former comparison. In reality, HANE, GRAPH and DART all share a location in another building which is far from NS, and JSCG and NS are next to each other.

One explanation for this could be that the buildings at locations NS and JSCG simply have a greater number of individual access points in them respectively when compared to other locations, and so the smartphone used to perform the hotspot scans has connected to the closest available hotspots, rather than those which are shared between both locations.

2.1.4 To Do 4

"Re-calculate and plot the cosine similarities with this combined data. Export and save these plot as Fig. 3 Similarity Data Combined. Comment on what you observe in these plots with regard to the results from a single person's data."

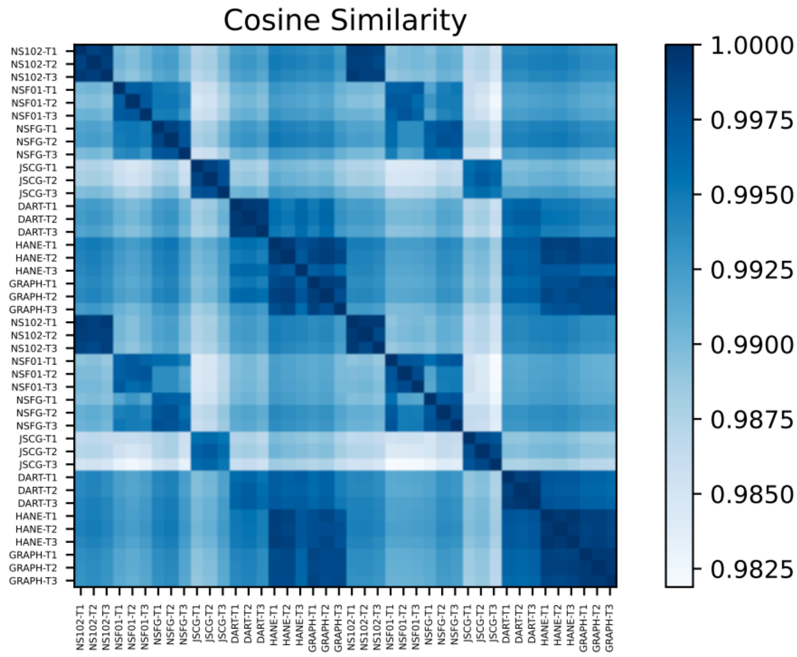

Figure 17: Cosine similarity matrix (combined data).

Figure 17 shows the cosine similarity matrix when additional data, collected by a second person, is added to the data pool. As with the first, the second data set contains WiFi access point scan data from the seven locations, with scans repeated in three trials. The first data set, D1, appears at the top of the y axis and the start of the x axis, and the second data set, D2, appears after them. The bottom left and top right quadrants show how similar the two data sets scan locations were.

In general, data sets D1 and D2 share similar patterns to each other. When comparing results from D2 to D1 cosine similarities, many of the suggestions made previously about geographical location from D1 hold true for D2 as well.

Interestingly the same inaccuracy that the D1 matrix showed in regards to the locations of HANE, GRAPH and DART in relation to JSCG and the location of NSC in relation to JSCG are clearer in D2. Looking at the lower right quadrant of the combined similarity matrix, D2's comparisons of the JSCG scans to other locations produce the similar "cross" pattern as D1 does. However, the shades of blue are on average much lighter and consistently lighter, indicating that not only is the inaccuracy present in D2, but is also inaccurate to a higher degree.

In addition, D2 shows, in general, smaller ranges in the cosine similarity values between trials at the same location. This results in more consistent colouration comparisons between trials at each location. This could possibly suggest that D2's trials were carried out within a smaller time window at each location than D1's, as the range of D2's accessed WiFi hotspots are smaller.

The smaller ranges in colouration also suggest that D2's trials were carried out in more similar positions at each location. However, the only location that does not follow this trend is the trials carried out at NS01 and NSG, in which the comparisons result in differing cosine similarity values. As mentioned previously, one explanation to this could be that the person who collected D2 took trials from varying points at the general locations of NS01 and NSG, not just the same physical spot.

Combining the cosine similarity matrices of D1 and D2 give a better visualization on which locations are geographically similar. They also give an understanding of how similar the scan locations are across trials which can therefore be used to better evaluate the validity of the data collection process and subsequent cosine similarity comparisons.

2.2 Stage 3 - User Feedback Design

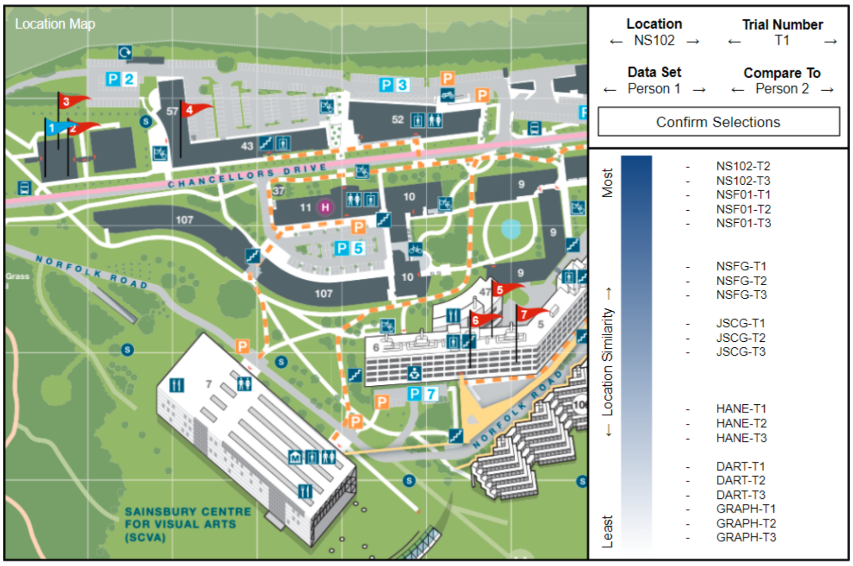

Figure 18: Proposed LoFi design.

Figure 18 shows a proposed LoFi design for representing the data from the cosine similarity matrix. The amount of data on the screen has been reduced in order for it to be more intuitive for a non-technical user to interpret. The design has been implemented with viewing on a standard 16:9 aspect ratio computer monitor in mind.

The screen window is divided into three sections. The most prominent section on the left shows a map of the area the locations occupy, magnified so that they all can appear simultaneously on screen. A blue flag replaces a red flag when the location has been selected on the right. Including this map of the locations allows the user to better understand how the data collected corresponds to the physical geographical locations. This allows easier comparisons so that the accuracy of the collected data can be analyzed.

The top right area of the window allows the user to select the location they want to look at, which person's data set to use, which location scan trial to use, and finally which data set to compare to. Table 3 shows the selection options for each category.

Selector	Options	Description
Location	See Table 1	Range of seven locations where WiFi access point scans were carried out.
Trial Number	T1, T2 or T3	Which scan trial at the location to select.
Data Set	Person 1 or Person 2	Which persons data set to take the trial from.
Compare To	Person 1, Person 2 or Both	Which data set or sets the trial will be compared to with.

Table 3: LoFi interface options.

The bottom right of the window shows the data from the cosine similarity matrix. Instead of displaying all comparisons at once, each location's trial is placed on an axis dependent on its cosine similarity value with the user selected trial. The higher the similarity value, the higher they appear on the list, and vice versa. The axis has been labeled with 'Location Similarity' to better communicate what the trial's ranking means. The colours from the cosine similarity matrix have been maintained, and each scan trial comparison is placed next to its corresponding colour.

Note that the data displayed in the LoFi design is not necessarily completely accurate, but added to demonstrate the purpose of the axis.

Overall the Lofi design attempts to reduce the amount of data presented to only what the user needs, potentially allowing for quicker understanding. This in addition to presenting the actual locations alongside the data's proposed similarities could potentially aid the user in evaluating which trials are accurate and which are inaccurate.