SHL dataset

SHL dataset documentation & QC

Prepared by: Dr. Daniel Roggen*, Dr. Hristijan Gjoreski*, Mathias Ciliberto*

*Sensor Technology Research Centre, School of Engineering and Informatics, University of Sussex, Brighton BN1 9QT, UK.

Date: 2 September 2017

Revised: 18 October 2017

Abstract

This deliverable comprises information about the data collection protocol as well as complete data collected during the project. The total recording time spanned 1084 hours with 3 users each carrying 4 phones on the hand, torso, hips and in a backpack. Out of the total recording time, 760 hours are fully annotated along 8 modes of locomotion and transportation as well as 5 additional tracks indicating: road type, traffic condition, presence in tunnels, social interaction and eating/drinking. If recordings containing irregularities identified due to reboots or alignment across phones are omitted entirely, the dataset comprises recordings spanning 700 annotated hours, where the data of all phones are available. The annotation of this data has been verified and corrected using our in-house annotation tool. Overall, the data is of high quality for all the movement sensors and pressure sensor as well as auxiliary sensors (battery, ambient light sensors, GPS). The Wifi and Cell information is of high quality in the second part of the dataset (after 19.04 for User 1, and 30.06 for User 2 and 3) while inadequate in the first part due to energy saving measures in Android. Google Activity recognition API output and location coverage are satisfactory.

1. Experimental setup

The experimental setup (fig. 1) comprises four Huawei Mate 9 phones and a body-worn camera for annotation. The phones are equipped with the data logging software explained in Deliverable 1.3.

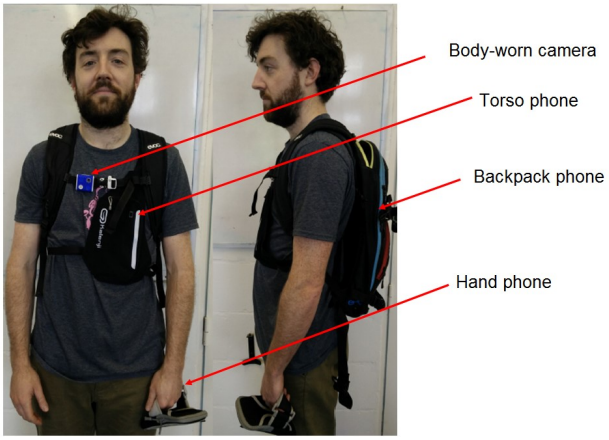
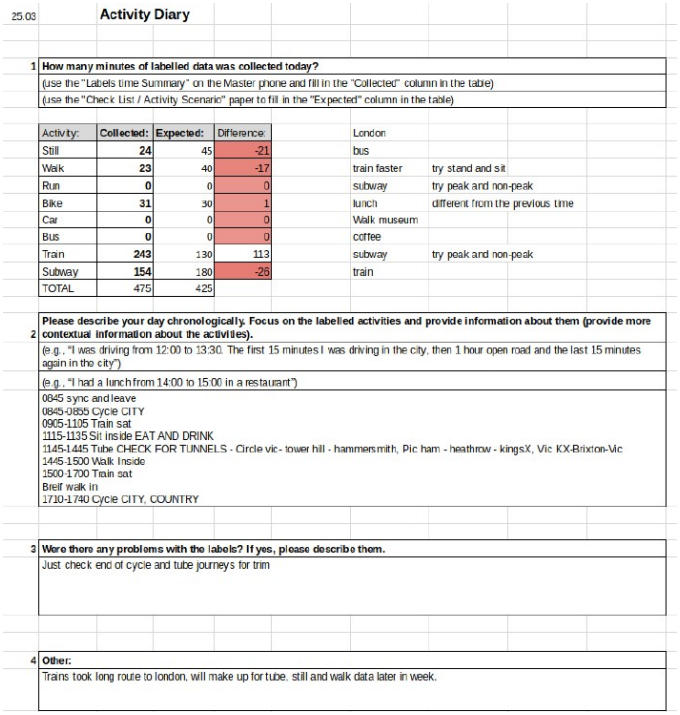


Figure 1. Experimental setup with four Huawei Mate 9 phones and a body-worn camera for annotation

The data collection plan is defined by the experiment leader and capabilities of the data collection subject following a weekly collection plan (fig. 2) available as a Google Sheet file. This plan comprises for each day of the week the envisioned transportation mode. During data collection, the user performs annotation directly on the app and writes notes (typically on a notebook and using the "flag" capability of the app) and reports in the collection plan the effective amount of data collected as well as notes relating to the data collection itself, such as clarifications about labels or potential issues.



25.03 **Activity Diary**

1 **How many minutes of labelled data was collected today?**
(use the "Labels time Summary" on the Master phone and fill in the "Collected" column in the table)
(use the "Check List / Activity Scenario" paper to fill in the "Expected" column in the table)

Activity:	Collected:	Expected:	Difference:
Still	24	45	-21
Walk	23	40	-17
Run	0	0	0
Bike	31	30	1
Car	0	0	0
Bus	0	0	0
Train	243	130	113
Subway	154	180	-26
TOTAL	475	425	

London		
bus		
train faster	try stand and sit	
subway	try peak and non-peak	
lunch	different from the previous time	
Walk museum		
coffee		
subway	try peak and non-peak	
train		

2 **Please describe your day chronologically. Focus on the labelled activities and provide information about them (provide more contextual information about the activities).**
(e.g., "I was driving from 12:00 to 13:30. The first 15 minutes I was driving in the city, then 1 hour open road and the last 15 minutes again in the city")
(e.g., "I had a lunch from 14:00 to 15:00 in a restaurant")

0845 sync and leave
0845-0855 Cycle CITY
0905-1105 Train sat
1115-1135 Sit inside EAT AND DRINK
1145-1445 Tube CHECK FOR TUNNELS - Circle vic - tower hill - hammersmith, Pic ham - heathrow - kingsX, Vic KX-Brixton-Vic
1445-1500 Walk Inside
1500-1700 Train sat
Breif walk in
1710-1740 Cycle CITY, COUNTRY

3 **Were there any problems with the labels? If yes, please describe them.**
Just check end of cycle and tube journeys for trim

4 **Other:**
Trains took long route to london, will make up for tube, still and walk data later in week.

Figure 2. Data collection plan as a Google Sheet, including the amount of expected data defined during setting up the protocol, a report of the effective amount of data collected, and additional notes relevant to the data collection or annotation.

The annotations comprises 8 primary categories, which are the objective of this project: Still, Walk, Run, Bike, Car, Bus, Train, and Subway. The project aims to have a balanced dataset among these 8 categories. These 8 categories are further broken down into: Still-Stand-Outside, Still-Stand-Inside, Still-Sit-Outside, Still-Sit-Inside, Walking-Outside, Walking-Inside, Run, Bike, Car, Bus-Stand, Bus-Sit, Train-Stand, Train-Sit, Subway-Stand, Subway-Sit.

One day a week, typically on Friday, the data collection subjects comes in the lab and the following procedure is followed:

1. The data of the phones is downloaded
2. The data of the body-worn camera is downloaded and the collection subject is left to remove unwanted privacy-sensitive pictures.
3. The body-worn camera pictures are transformed in a timelapse video and imported in our in-house annotation tool with the phone data.
4. The collection subject corrects possible issues with annotations and provides additional annotations of relevance which are not suitable to be collected directly on the phone

5. The raw data are converted in an intermediate "merged" representation with the annotation.
6. The raw data and the merged representation are backed up on- and off-site.

2. Quality control

The merged representation is further processed to extract statistics about the data (see appendix) which inform our quality control.

2.1 Manual quality control

Some of the quality control is done through visual inspection of time-interval histograms and data distribution.

Time-interval histograms are especially useful for sensors which have a regular sample rate, which on the Mate 9 include all the motion sensors and the pressure sensor. The firmware of the phones of User 1 were updated on the 11 May 2017 and the sample rate of the motion and pressure sensor changed due to the firmware update. Prior to the updated and for the phones of User 2 and 3, all the motion sensors and the pressure sensors operate at 100Hz. This is shown in the time interval histogram with a clear peak at 10ms (fig. 3). In figure 3 and later, "Sens" refers to the sensor time which is available in the Android sensor event; most Android implementations store the time when the sample was acquired in this field. "Host" refers to the time when the Android logging application receives the sample from the operating system. While older hardware often displayed significant differences between the sensor and host time, we did not notice significant differences between the two reported times. Note that the Sensor time has a granularity of 1 ns while the host time has a granularity of 1 ms, which explains the slightly different histogram shapes.

After the firmware update of the phones of User 1 on 11 May 2017, the acceleration and gyroscope sensors are sampled at 200Hz or 250Hz with the screen off and on respectively, the pressure sensor is sampled at 10Hz, and the other motion sensors are sampled at 100Hz. We observed significantly more jitter in the time intervals due to the higher sample rate and the variation between 200Hz and 250Hz.

Other sensors have a more irregular sample rate defined by the low-level Android driver implementation. This includes battery (sample rate about 1 per minute), light sensor (about 3 Hz), Google recognition API (sample rate experimentally to be about once every 10 second, although the Google API reports that samples are provided whenever the activity changes) and location (sample rate about 1Hz) and GPS (sample rate about 1Hz) (fig. 4).

We also visualise a histogram of the sensor data itself (fig. 5 and fig. 6). This allows to spot possible data outliers caused either by errors in the application logging code, or potentially in the Android sensor driver. Once we verified that there are no errors in the logging code this visualisation was not useful anymore. In particular we never observed unexpected sensor readouts.

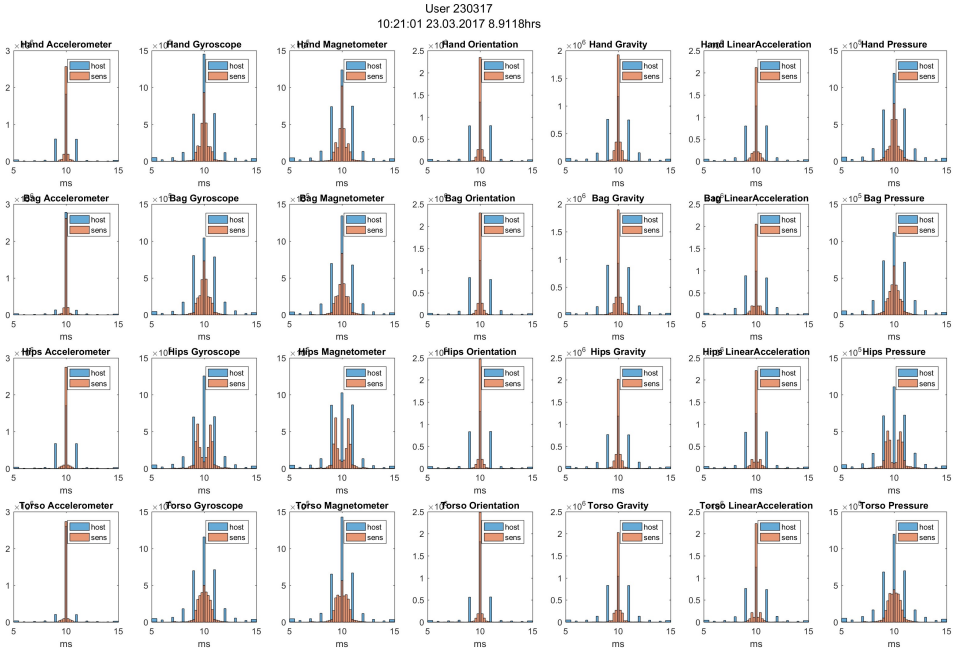


Figure 3. Sample time interval for the motion sensors and pressure sensor showing a clear peak at 10 ms (100Hz sample rate).

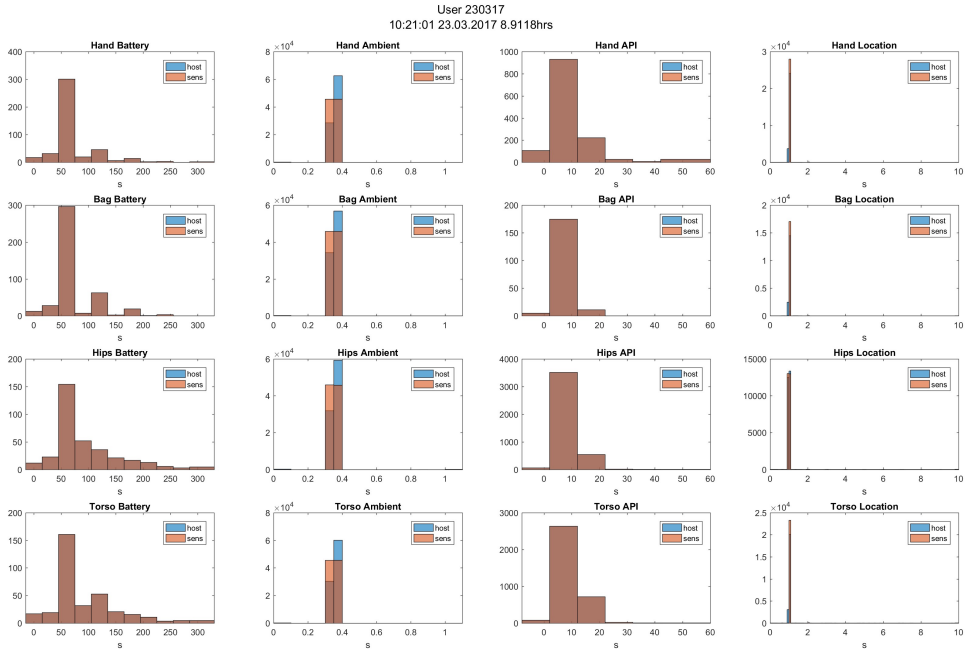


Figure 4. Sample time interval for the battery (expected 1/min), ambient light sensor (expected about 3Hz), Google API (experimentally observed to be about once every 10 seconds although the Android documentation reports a variable sample rate given by user activity transitions) and the location sensor (expected 1Hz).

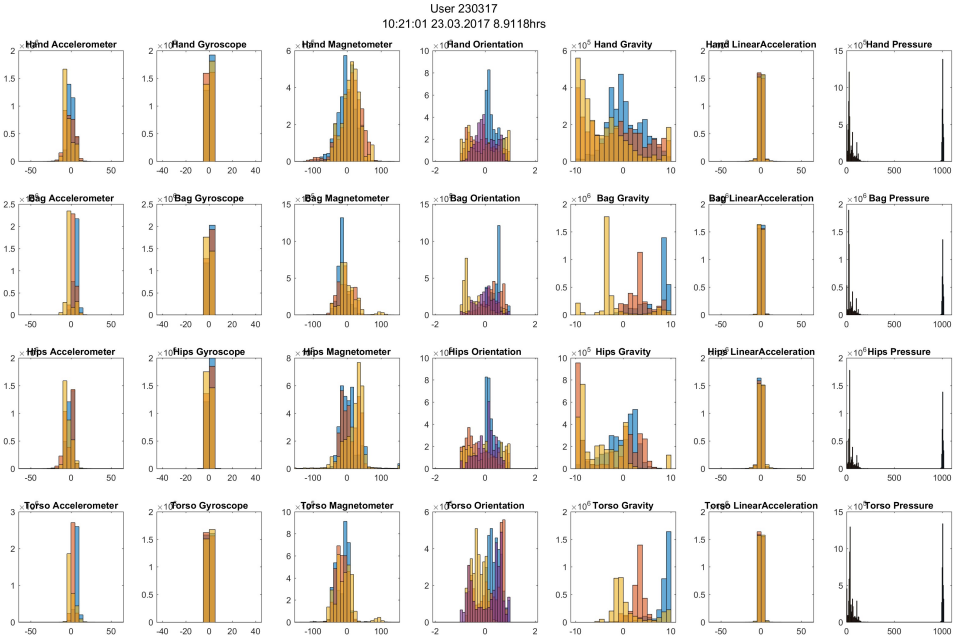


Figure 5. Histogram of motion and pressure sensor data. All 3 motion channels (or 4 for the quaternion orientation) are expected to have overlapping distributions in a physically realistic range. The pressure sensor provides pressure (peak at about 1000 mBar), temperature and altitude difference (peaks around 20 and 0).

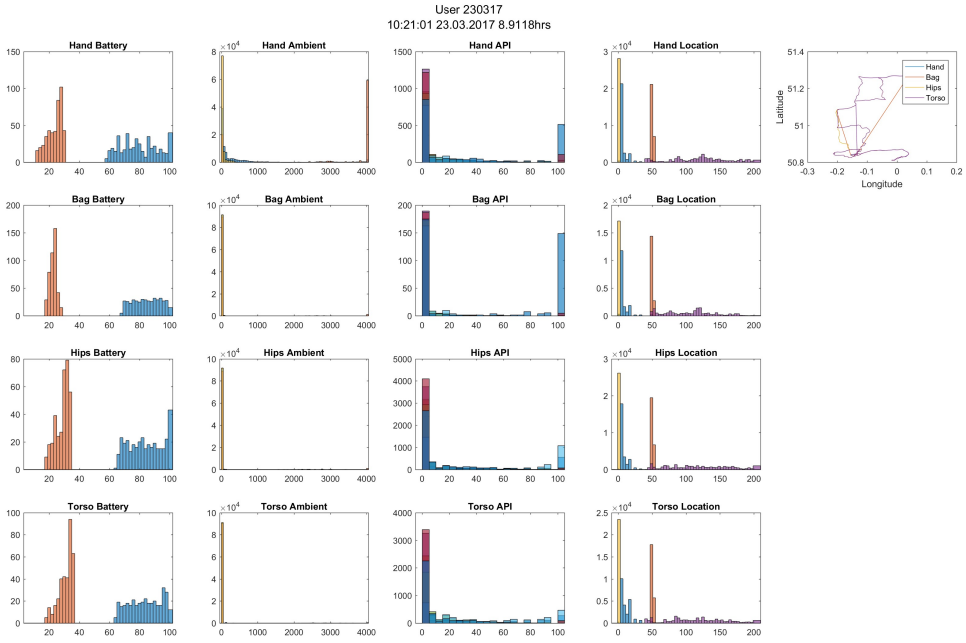


Figure 6. Histogram of sensor data for battery, ambient light, Google recognition API and location, including location trace (top right). We expect battery to show two distributions: one from 50 to 100 representing charge and one around 20 representing the temperature. Location should show two

distributions around 0 and 50 representing the latitude/longitude of the location where the recordings take place. The ambient light should show distributions around 20 (temperature) and 4000 (light temperature). The distribution of the Google recognition API should be between 0 and 100.

Some sensors provide highly irregular data and/or variable number of data channels: this includes the cell scan and Wifi scan. We visualised time interval histogram as well as time interval boxplots and overall number of samples acquired (fig. 7). Figure 7 also shows the GPS time interval histogram and GPS sensor data histogram. We found that for sensors with a highly irregular data sampling (Wifi, cells) the time interval histogram is not useful and a better visualisation is a "coverage map".

The coverage map (fig. 8) represents the amount of data received per one minute for each sensor channels, normalised by the expected number of data experimentally determined from test recordings. The 0 of samples). Overall, a good quality recording would show bars of identical height for each channels at regular intervals. Missing bars or bars of comparatively lower height for one channel indicate decrease in sample rate. This can happen without being a sign of error for GPS and Location (if no satellites are visible) and for the Google API (if the user does not transition between activities). Figure 8 illustrates a good quality recording with a consistent coverage of sensor sampling throughout the recording.

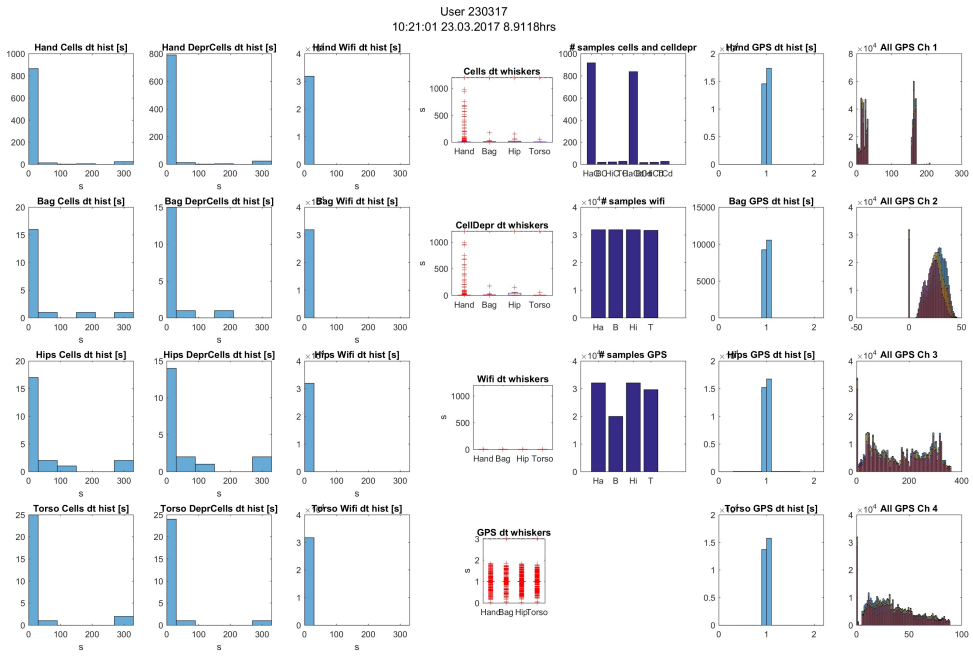


Figure 7. Histogram of time intervals and boxplot of time intervals for cells and Wifi (left 4 columns); number of samples for cells, Wifi and GPS (fifth column), and time interval histogram for GPS (expected 1 second, second column to the right) and GPS sensor data histogram (rightmost column).

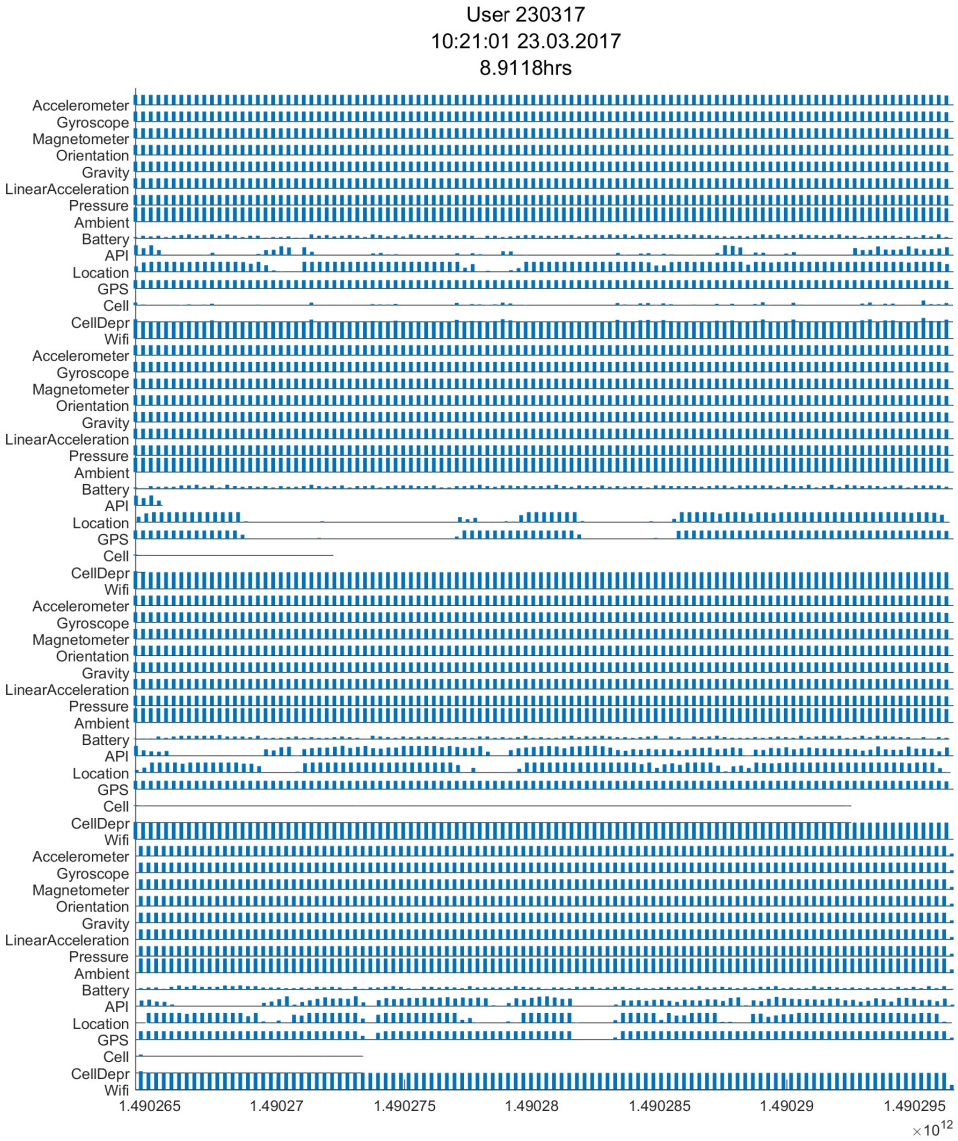


Figure 8. Coverage map illustrating a high quality coverage for all the motion sensors, the pressure sensor, the battery sensor (which has a variable sample rate), and the Wifi scans. Note that location and GPS are highly correlated and some phone positions offer better GPS/location coverage than others. Cells and the deprecated cell API have missing coverage due to the energy saving features of Android which scans cells only when deemed necessary. The Google location API provides highly irregular coverage as it reports data on changes of activities only.

2.2 Automatic quality control

In addition to the visual inspection we created quantitative automatic checks. These include: primarily the assessment of a continuous "coverage metric". This value is 1 if the expected number of samples is obtained. It is higher or lower than 1 if respectively more or less samples than expected are received. This coverage metrics works reliably for sensors providing regular samples, which includes the motion sensors, pressure sensor, ambient light and battery. Other sensors provide regular samples when they are in the condition to do so: e.g. location provides regular samples if a GPS lock is available but would not provide samples otherwise (e.g. indoors or in a tunnel).

The coverage metric is the ratio between the received number of samples and expected number of samples. This can be easily computed for sensors with a regular sample rate (all the motion sensors and the pressure sensors at 100Hz). However to make the system more adaptive to the different channels we automatically computed the expected sample rate using the median of time intervals. Therefore, the coverage metric is the total number of samples received divided by the expected number of samples, which is the duration of the recording divided by the median of the time intervals. We found this metric to be satisfactory for all sensors.

Additionally, we check the alignment of start/end time of all channels, where the first and last data of all channels must be within a narrow time window (15mn). We also automatically detect possible reboots of a phone. As the application restarts logging automatically if a reboot occurs this measure is used to identify if a user manipulation error occurred (e.g. long-pressing the power button leading to a reboot).

Finally we converted these metrics into a binary decision about the quality of each sensor channels for each phone.

All motion sensors and pressure (which are all sampled regularly at 100Hz or 200/250Hz) must each have a coverage between 92% to 108% using the above metric. The ambient light sensors must have coverage between 95% and 105%, the battery must have coverage >60%, the GPS[1] must have coverage > 75%, the Google recognition API must have coverage > 50% (the lower value is related to this API providing information only on transitions between activities), the location must have a coverage >50% (this is lower than the GPS coverage as the satellites must be visible with good enough signal quality for long enough to deliver a coordinate) and cells and Wifi must have coverage >75%. In addition to the coverage, all the sensor channels must have the first/last sample within a narrow time window of each other, otherwise the channel is declared as not good.

We report this information in a condensed manner by grouping all the motion sensors and the pressure sensor together yielding one binary measure of quality. Battery, GPS and ambient light are combined in a single binary measure of quality (referred to as "auxiliary sensors" in the report). Location, cells and Wifi are individually reported.

Figure 9 explains the overall summary table.

Rec	Start	Len	Reboot	M+P	Aux	Loc	Cell	Wifi
010317	13:57:02 01.03.2017	4.3161hrs	0 0 0 0	1.0 1.0 1.0 1.0 1 1 1 1	0.9 0.9 0.9 0.9 1 1 1 1	0.7 0.7 0.6 0.6 1 1 1 1	0.2 0.0 0.0 0.0 0 0 0 0	0.8 0.8 0.8 0.8 0 0 0 0

Figure 9. Summary of quality control with the record ID, start time and length of recording (left 3 columns), the number of reboots for each of the 4 phones in the order hand, bag, hip, torso. The M+P column represents all the movement sensors and pressure sensor. The first line indicates the coverage (1 is maximal) and the second line is the binary decision if the channel are ok. Aux refers to battery, GPS and ambient light and indicates the aggregate coverage for these 3 sensors and whether they are all ok. The remaining three columns report coverage and ok for location, cells and Wifi.

Automatic quality control results are marked in the tables in appendix and following sections with a symbol in the margin, such as ✓.

[1] "GPS" is the sensor indicating the azimuth/elevation and SNR of visible satellites. The sensor providing the device coordinates is the "Location" sensor.

✓ is used to indicate recordings where no phone rebooted during the recording, where the number of motion and pressure samples is within tolerance.

✓+L is used to indicate recordings where in addition the number of location samples is within tolerance. Note that if the recording contains significant portion of time underground or indoors there may be a lower than expected number of location samples and the recording would not be flagged by ✓+L although this does not indicate an issue.

✓+W and ✓+C is used to indicate recordings were in addition the number of Wifi and Cell samples are within tolerance.

2.3 Summary of quality control

We recommend to only employ recordings flagged with ✓ for any data analysis. For such recordings the motion and pressure sensor data is of high quality.

Wifi scanning and Cell scanning was insufficient during some of the early recordings due to relying on a "best effort" scanning of Android. We then changed the logging software to force reporting of Wifi and Cell at 1 second interval. While this forces the operating system to issue sensor events to the application at that rate, Android does not guarantee that this changes the underlying driver Wifi and cells scan rate.

In general we have not observed any issue with the sensing of GPS (i.e. satellite azimuth, elevation and SNR).

Location coverage is function of the user's location and may be low if indoors or in subways. In several cases we observed no location reported on some of the phones but not on the others during extended periods (e.g. during train rides).

The Google activity recognition API coverage is clearly linked to availability of location information. The sampling is also irregular as the Google activity recognition API seems to have a heuristic to report activity changes only sporadically, potentially to minimise power usage.

Coverage of the battery and ambient light sensors is often low, however these are not modalities relevant for activity recognition.

3. Dataset summary

3.1 Terminology

The dataset comprises *users* providing data in multiple *recordings*. Each *recording* corresponds to an uninterrupted data collection, generally comprising a half day up to a full day of data.

The complete dataset comprises 3 *users* (User 1, User 2 and User 3). The data of User 1 corresponds to the data collected during the preliminary data collection phase (which was covered in D1.1) and the follow-up data collection phase. As the same person was recorded both in the preliminary data collection phase and in the follow-up we combine the data

Table 1. Sex of participants

User	Collected during	Sex
1	Preliminary + follow-up	Male
2	Follow-up	Male
3	Follow-up	Female

3.2 Summary tables

The following tables report the number of annotated data for the 3 recording session (User1, User2, User3), each corresponding to a unique person. The recording sessions User1 combines the data of the preliminary data collection reported in D1.1 (which was called 'User' in that deliverable) and the additional data collected with the same user in the follow-up data collection phase.

In the following tables, "Total" indicates the total duration of activities which was annotated. Total marked with ✓ is the total amount of annotated data for all recordings where not a single phone experienced motion sensor data loss[2] during the recording. In other words, Total✓ indicates the total amount of data where the user's 4 phones provide motion data throughout the recording. This measure is a very conservative indication of the amount of data available, as any reboot of any one phone during one recording leads to discarding the entire recordings comprising the data of the 4 phones.

Total marked with ✓+L indicates the subset of data reported in Total✓, where in addition the location sensor provides sufficient coverage throughout the recording according to the automated quality checks. This measure is highly conservative, as it is normal that location is not available in certain situations, such as in subway or indoor environments and we observed recordings where some of the phones take a long time to acquire location data.

Total marked with ✓+C and ✓+W indicates the subset of data reported in Total✓ where, in addition the Cellular data and Wifi data respectively provides sufficient coverage throughout the recording according to the automated quality checks.

[2] Data loss was either due to sporadic reboot, or to a user inadvertently turning off a phone by long-pressing the power button.

Table 2. Amount of annotated data in hours collected with user 1.

User 1 [hrs]	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Total
Total	57.21	51.75	17.68	48.32	68.24	51.1	56.25	40.45	391
Total✓	54.31	48.19	17.07	43.52	62.58	49.61	52.75	38.21	366.25
Total✓+L	23.05	24.68	9.28	19.25	18.78	30.43	10.93	2.57	138.95
Total✓+C	32.93	33.03	11.09	30.43	46.84	38.31	38.58	30.09	261.3
Total✓+W	41.81	38.13	13.09	36.98	54.17	42.27	48.84	35.68	310.97

Table 3. Amount of annotated data in hours collected with user 2.

User 2 [hrs]	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Total
Total	32.22	41.96	0.75	19.46	29.64	29.66	28.62	22.58	204.89
Total✓	30.32	38.12	0.5	19.46	29.53	28.98	24.54	19.77	191.22
Total✓+L	26.88	32.29	0.5	19.46	29.47	26.68	12.24	4.72	152.26
Total✓+C	19.59	23.71	0	15.11	21.89	20.45	14.14	15.05	129.94
Total✓+W	19.59	23.71	0	15.11	21.89	20.45	14.14	15.05	129.94

Table 4. Amount of annotated data in hours collected with user 3

User 3 [hrs]	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Total
Total	34.71	29.3	4.58	18.99	1.4	26.47	27.22	23.48	166.14
Total✓	29.73	26	4.06	17.09	0	24.49	23.15	20.21	144.74
Total✓+L	7.92	7.76	1.02	5.19	0	12.65	1.7	0	36.23
Total✓+C	16.72	13.81	2.26	10.03	0	6.17	13.8	17.3	80.09
Total✓+W	16.72	13.81	2.26	10.03	0	6.17	13.8	17.3	80.09

Table 5. Grand total annotated data in hours collected with all users

Overall [hrs]	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Total
Total	124.14	123.01	23.01	86.77	99.28	107.23	112.09	86.51	762.03
Total ✓	114.36	112.31	21.63	80.07	92.11	103.08	100.44	78.19	702.21
Total ✓+L	57.85	64.73	10.8	43.9	48.25	69.76	24.87	7.29	327.44
Total ✓+C	69.24	70.55	13.35	55.57	68.73	64.93	66.52	62.44	471.33
Total ✓+W	78.12	75.65	15.35	62.12	76.06	68.89	76.78	68.03	521

3.2 Individual recording quality and duration

The following tables report the automatic evaluation of the quality of each recording. The leftmost column indicates the automatic quality evaluation. A missing ✓ indicates that one or more phones rebooted during the recording and/or that there is insufficient data obtained from the motion sensors and/or that there is inconsistent recording start and end time. The symbols L, C, W next to ✓ indicate that there is sufficient coverage from the location, cells and Wifi sensors, respectively. It is recommended to discard recordings which are not at minimum affixed by ✓ from any data analysis.

Table 6. User 1 recording quality and duration, for the data collected during preliminary data collection phase.

	Rec	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Tot
✓	m010317	0.56	0.36	0.00	0.37	0.00	0.00	0.00	0.00	1.29
✓$\mathcal{L}$	010317	0.78	1.94	0.00	0.29	0.00	0.00	0.00	0.00	3.01
✓$\mathcal{L}$	020317	3.26	1.90	0.53	0.00	0.00	2.27	0.00	0.00	7.96
✓$\mathcal{L}$	m030317	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.25
✓	030317	1.09	0.10	0.00	0.45	0.00	0.17	1.28	0.00	3.09
✓$\mathcal{L}$	060317	0.29	0.29	0.40	0.00	0.00	0.91	0.00	0.00	1.90
	070317	1.14	1.10	0.00	0.00	0.00	0.49	2.05	2.25	7.03
✓	080317	1.25	0.51	1.00	2.09	0.11	0.00	0.00	0.00	4.96
✓$\mathcal{L}$	090317	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.43
✓$\mathcal{L}$	100317	0.20	0.16	0.48	0.00	0.00	0.65	0.00	0.00	1.49
✓$\mathcal{L}$	130317	0.95	1.27	0.59	0.67	0.00	2.57	0.00	0.00	6.06
✓	140317	1.14	1.28	0.00	0.20	0.00	0.48	2.63	2.53	8.26
✓$\mathcal{L}$	150317	2.08	1.75	0.28	0.00	4.54	0.10	0.00	0.00	8.74
✓	160317	0.85	0.45	0.00	1.78	3.77	0.00	0.00	0.00	6.85
✓$\mathcal{L}$	m170317	0.07	0.03	0.50	0.00	0.00	0.00	0.00	0.00	0.60
✓$\mathcal{L}$	170317	0.00	0.00	0.19	0.00	0.00	0.19	0.00	0.00	0.38
✓$\mathcal{L}$ $\mathcal{W}$	210317	1.51	0.97	0.00	1.90	0.00	0.00	2.35	0.00	6.72
✓$\mathcal{W}$	220317	2.57	1.06	0.53	0.12	3.95	0.00	0.00	0.00	8.23
✓$\mathcal{L}$ $\mathcal{W}$	230317	0.52	0.00	0.00	0.97	3.38	0.00	0.60	0.00	5.47
✓$\mathcal{L}$ $\mathcal{W}$	m240317	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	0.45
✓$\mathcal{W}$	240317	0.92	0.14	0.00	0.00	0.00	0.35	0.00	0.00	1.41
✓$\mathcal{W}$	250317	0.43	0.33	0.00	0.50	0.00	0.00	4.02	2.55	7.83
✓$\mathcal{L}$ $\mathcal{W}$	270317	1.37	0.79	0.52	2.18	0.00	0.00	1.17	0.00	6.04
✓$\mathcal{W}$	280317	0.92	1.18	0.00	0.00	0.00	0.31	2.12	3.04	7.58
✓$\mathcal{L}$ $\mathcal{W}$	m290317	0.09	0.08	0.00	0.00	0.00	0.85	0.00	0.00	1.02
✓$\mathcal{L}$ $\mathcal{W}$	290317	0.55	0.55	0.50	0.87	0.00	2.10	0.00	0.00	4.59
✓$\mathcal{L}$ $\mathcal{W}$	300317	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.34
	Total	22.52	16.26	5.98	13.09	15.74	11.80	16.22	10.36	111.98
	Total ✓	21.38	15.16	5.98	13.09	15.74	11.30	14.17	8.12	104.95
	Total ✓$\mathcal{L}$	11.66	9.75	4.46	7.58	7.91	9.99	4.12	0.00	55.46
	Total ✓$\mathcal{C}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Total ✓$\mathcal{W}$	8.88	5.10	2.00	6.55	7.33	3.96	10.26	5.59	49.67

Table 7. User 1 recording quality and duration, for the data collected during the follow-up data collection phase.

	Rec	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Tot
✓$\mathcal{C}$ $\mathcal{W}$	190417	0.97	1.05	0.51	2.30	0.00	0.00	0.94	0.00	5.77
✓$\mathcal{C}$ $\mathcal{W}$	200417	0.18	0.90	0.00	0.77	0.00	0.00	2.49	2.55	6.89
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m240417	0.00	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.23
✓$\mathcal{C}$ $\mathcal{W}$	240417	0.17	0.13	0.00	0.52	0.00	0.97	0.00	0.00	1.79
✓$\mathcal{C}$ $\mathcal{W}$	250417	0.07	0.19	0.00	0.00	6.54	0.18	0.00	0.00	6.97
✓$\mathcal{C}$ $\mathcal{W}$	260417	0.60	0.70	0.25	1.64	3.14	0.00	0.00	0.00	6.33
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	270417	1.54	0.63	0.52	0.32	0.20	2.95	1.48	0.00	7.64
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m280417	0.00	0.03	0.00	0.00	0.00	0.31	0.00	0.00	0.35
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	280417	0.05	0.24	0.00	0.00	0.00	0.31	1.01	0.00	1.60
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	020517	0.48	0.84	0.00	0.00	0.27	0.00	2.34	2.57	6.50
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	030517	1.02	1.00	0.52	0.00	0.00	3.36	0.00	0.00	5.90
✓$\mathcal{C}$ $\mathcal{W}$	040517	0.49	0.38	0.00	0.00	0.00	0.57	2.31	3.04	6.79
✓$\mathcal{C}$ $\mathcal{W}$	050517	1.34	0.55	0.54	1.84	0.00	0.00	2.04	0.00	6.30
✓$\mathcal{C}$ $\mathcal{W}$	m080517	0.00	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.77
	080517	0.00	0.00	0.00	0.30	1.74	0.00	0.00	0.00	2.04
✓$\mathcal{C}$ $\mathcal{W}$	090517	0.55	1.00	0.00	0.00	0.00	0.48	2.45	2.48	6.97
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	100517	0.91	1.05	0.53	0.89	0.00	2.96	0.00	0.00	6.34
✓$\mathcal{C}$ $\mathcal{W}$	110517	0.46	0.13	0.16	1.95	4.34	0.00	0.00	0.00	7.03
✓$\mathcal{C}$ $\mathcal{W}$	120517	0.92	2.20	0.55	0.05	2.90	0.00	0.00	0.00	6.62
✓$\mathcal{C}$ $\mathcal{W}$	m150517	0.27	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.53
✓$\mathcal{C}$ $\mathcal{W}$	150517	0.52	0.11	0.00	0.40	0.00	0.00	0.00	0.00	1.02
✓$\mathcal{C}$ $\mathcal{W}$	170517	0.55	1.04	0.00	0.00	0.00	0.58	2.74	2.41	7.33
✓$\mathcal{C}$ $\mathcal{W}$	190517	0.63	0.48	0.62	0.50	5.35	1.64	0.00	0.00	9.22
✓$\mathcal{C}$ $\mathcal{W}$	200517	0.29	0.00	0.00	0.00	4.42	0.00	0.00	0.00	4.70
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m220517	0.00	0.09	0.00	0.83	0.00	0.00	0.00	0.00	0.92
✓$\mathcal{C}$ $\mathcal{W}$	220517	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.32
✓$\mathcal{C}$ $\mathcal{W}$	230517	2.74	0.24	0.00	0.00	0.00	3.11	0.00	0.00	6.09
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	240517	0.67	2.64	0.00	0.00	0.16	2.41	0.00	0.00	5.88
✓$\mathcal{C}$ $\mathcal{W}$	250517	0.21	0.65	0.25	0.90	0.00	0.00	2.57	2.57	7.15
✓$\mathcal{C}$ $\mathcal{W}$	260517	0.59	0.00	0.00	3.10	0.00	0.00	1.70	0.00	5.38
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	290517	0.37	0.82	0.52	0.00	0.00	0.00	0.00	0.00	1.72
✓$\mathcal{C}$ $\mathcal{W}$	m300517	0.00	0.00	0.00	0.00	2.03	0.00	0.00	0.00	2.03
✓$\mathcal{C}$ $\mathcal{W}$	300517	0.60	0.00	0.60	0.00	0.30	0.50	0.00	0.00	2.00
✓$\mathcal{C}$ $\mathcal{W}$	310517	0.88	1.21	0.30	0.00	0.00	0.00	2.43	2.42	7.23
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	010617	0.87	1.58	0.50	0.00	3.96	0.18	0.00	0.00	7.10
	020617	0.92	0.62	0.31	2.03	3.05	0.00	0.00	0.00	6.93
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	030617	0.64	0.47	0.55	1.75	0.08	1.97	0.00	0.00	5.45
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m050617	0.18	0.08	0.57	0.00	0.00	0.24	0.00	0.00	1.08
	050617	0.10	0.37	0.00	0.00	0.00	0.16	0.00	0.00	0.63
✓$\mathcal{C}$ $\mathcal{W}$	060617	0.78	0.85	0.00	0.00	0.00	0.00	2.27	2.96	6.86
✓$\mathcal{C}$ $\mathcal{W}$	070617	0.34	0.97	0.00	0.20	0.00	3.41	0.00	0.00	4.91
	080617	0.71	0.95	0.25	2.46	0.87	0.00	1.44	0.00	6.68
✓$\mathcal{C}$ $\mathcal{W}$	m090617	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.50
	090617	0.03	0.12	0.05	0.00	0.00	0.35	0.00	0.00	0.55
✓$\mathcal{C}$ $\mathcal{W}$	120617	0.78	0.41	0.00	0.28	0.00	0.00	2.59	2.43	6.49
✓$\mathcal{C}$ $\mathcal{W}$	130617	0.81	0.33	0.00	0.19	0.00	4.68	0.00	1.61	7.62
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	140617	0.60	1.21	0.00	1.80	0.24	0.72	0.00	0.00	4.56
✓$\mathcal{C}$ $\mathcal{W}$	150617	1.07	0.00	0.00	0.00	0.00	0.39	0.00	0.00	1.46
✓$\mathcal{C}$ $\mathcal{W}$	200617	0.84	1.33	0.80	1.11	0.75	0.00	1.73	0.00	6.56
✓$\mathcal{C}$ $\mathcal{W}$	220617	0.70	1.08	0.34	0.00	2.76	0.20	0.00	0.00	5.07
✓$\mathcal{C}$ $\mathcal{W}$	230617	0.61	1.06	0.17	0.78	3.45	0.00	0.00	0.00	6.06
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	260617	1.18	1.00	0.03	1.04	0.00	2.54	0.00	0.00	5.79
✓$\mathcal{C}$ $\mathcal{W}$	270617	0.54	0.15	0.00	0.34	0.00	0.00	2.49	2.53	6.05
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	280617	1.31	0.50	0.00	2.44	0.76	0.00	1.98	0.00	7.00
✓$\mathcal{C}$ $\mathcal{W}$	m290617	0.00	0.09	0.58	0.00	0.00	0.32	0.00	0.00	0.99
✓$\mathcal{C}$ $\mathcal{W}$	290617	1.41	0.09	0.00	0.55	0.00	0.49	0.00	0.00	2.55
✓$\mathcal{C}$ $\mathcal{W}$	300617	0.65	0.78	0.11	0.00	0.00	0.36	3.03	2.52	7.46
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	030717	0.97	1.08	0.50	0.85	0.00	2.50	0.00	0.00	5.89
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	040717	0.07	0.45	0.00	1.53	3.08	0.00	0.00	0.00	5.14
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	050717	0.52	1.21	0.57	0.00	2.12	0.00	0.00	0.00	4.42
	m060717	0.00	0.40	0.00	0.00	0.00	0.48	0.00	0.00	0.88
	Total	34.69	35.49	11.70	35.23	52.50	39.30	40.03	30.09	279.02
	Total ✓	32.93	33.03	11.09	30.43	46.84	38.31	38.58	30.09	261.30
	Total ✓$\mathcal{L}$	11.39	14.93	4.82	11.67	10.87	20.44	6.81	2.57	83.49
	Total ✓$\mathcal{C}$	32.93	33.03	11.09	30.43	46.84	38.31	38.58	30.09	261.30
	Total ✓$\mathcal{W}$	32.93	33.03	11.09	30.43	46.84	38.31	38.58	30.09	261.30

Table 8. User 2 recording quality and duration.

	Rec	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Tot
✓ $\mathcal{L}$	070617	0.93	0.90	0.00	0.00	0.00	0.98	0.00	0.00	2.80
	080617	1.49	2.86	0.25	0.00	0.00	0.68	1.35	0.00	6.63
✓ $\mathcal{L}$	m090617	0.10	0.38	0.00	0.00	0.00	0.22	0.00	0.00	0.71
✓ $\mathcal{L}$	120617	0.81	2.16	0.25	0.00	0.00	0.66	2.80	0.00	6.68
✓ $\mathcal{L}$	m130617	0.06	0.35	0.00	0.00	0.00	0.25	0.00	0.00	0.66
✓ $\mathcal{L}$	130617	1.47	0.41	0.00	0.00	3.99	0.29	0.00	0.00	6.16
✓ $\mathcal{L}$	140617	1.57	2.24	0.25	2.27	0.00	0.61	0.00	0.00	6.95
✓ $\mathcal{L}$	m150617	0.03	0.36	0.00	0.00	0.00	0.20	0.00	0.00	0.59
✓ $\mathcal{L}$	160617	1.38	2.26	0.00	0.00	0.00	0.59	2.08	0.00	6.32
✓ $\mathcal{L}$	190617	0.70	1.22	0.00	0.32	0.00	0.38	3.01	2.23	7.85
✓ $\mathcal{L}$	200617	1.00	1.67	0.00	0.00	0.00	3.59	0.00	0.00	6.26
✓ $\mathcal{L}$	m210617	0.00	0.33	0.00	0.00	0.00	0.32	0.00	0.00	0.66
✓ $\mathcal{L}$	220617	1.92	0.96	0.00	0.99	3.65	0.00	0.00	0.00	7.52
✓ $\mathcal{L}$	260617	0.32	0.90	0.00	0.00	0.00	0.28	2.51	2.49	6.50
✓ $\mathcal{L}$	m280617	0.45	0.26	0.00	0.77	0.00	0.16	0.00	0.00	1.64
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	280617	0.12	1.04	0.00	0.54	0.00	0.28	0.00	0.00	1.98
✓ $\mathcal{C}$ $\mathcal{W}$	300617	0.76	1.35	0.00	0.00	0.00	0.38	1.90	3.00	7.39
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	010717	1.01	1.02	0.00	0.00	0.00	1.52	0.00	0.00	3.55
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	020717	0.51	0.50	0.00	2.01	0.00	0.00	0.00	0.00	3.02
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m030717	0.39	0.34	0.00	0.00	0.00	0.23	0.00	0.00	0.96
✓ $\mathcal{C}$ $\mathcal{W}$	040717	0.60	1.16	0.00	0.00	0.00	0.47	2.35	3.00	7.59
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	060717	3.07	1.48	0.00	2.00	0.00	1.99	0.00	0.00	8.53
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	070717	0.80	2.37	0.00	0.00	0.00	1.17	1.84	0.00	6.18
✓ $\mathcal{C}$ $\mathcal{W}$	080717	0.57	0.50	0.00	0.00	0.06	0.20	1.58	0.13	3.04
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	120717	1.29	1.81	0.00	0.50	0.00	2.51	0.00	0.00	6.12
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m130717	0.00	0.30	0.00	0.00	0.00	0.22	0.00	0.00	0.53
✓ $\mathcal{C}$ $\mathcal{W}$	140717	0.64	1.13	0.00	0.00	0.00	0.28	2.24	2.95	7.24
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	160717	1.35	1.71	0.00	1.04	0.00	2.02	0.00	0.00	6.11
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	170717	1.00	0.92	0.00	1.96	0.00	2.07	0.00	0.00	5.95
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	180717	1.17	0.76	0.00	0.00	4.29	0.27	0.00	0.00	6.50
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	190717	0.48	0.66	0.00	0.00	4.85	0.14	0.00	0.00	6.13
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	200717	1.05	1.02	0.00	2.00	0.00	1.99	0.00	0.00	6.06
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m210717	0.06	1.30	0.00	0.00	0.00	0.52	0.00	0.00	1.88
	240717	0.41	0.98	0.00	0.00	0.10	0.00	2.73	2.81	7.03
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	250717	1.01	0.98	0.00	2.02	0.00	2.01	0.00	0.00	6.02
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	270717	1.10	0.58	0.00	0.00	4.30	0.19	0.00	0.00	6.17
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	280717	1.23	0.31	0.00	0.00	4.39	0.50	0.00	0.00	6.42
✓ $\mathcal{C}$ $\mathcal{W}$	310717	0.80	1.23	0.00	0.00	0.00	0.19	2.09	2.76	7.07
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m010817	0.02	0.33	0.00	0.00	0.00	0.22	0.00	0.00	0.57
✓ $\mathcal{C}$ $\mathcal{W}$	020817	0.06	0.47	0.00	0.00	0.00	0.76	2.14	3.22	6.65
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	030817	0.42	0.47	0.00	1.12	4.00	0.31	0.00	0.00	6.32
✓ $\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	m040817	0.08	0.00	0.00	1.91	0.00	0.00	0.00	0.00	1.99
	Total	32.22	41.96	0.75	19.46	29.64	29.66	28.62	22.58	204.89
	Total ✓	30.32	38.12	0.50	19.46	29.53	28.98	24.54	19.77	191.22
	Total ✓ $\mathcal{L}$	26.88	32.29	0.50	19.46	29.47	26.68	12.24	4.72	152.26
	Total ✓ $\mathcal{C}$	19.59	23.71	0.00	15.11	21.89	20.45	14.14	15.05	129.94
	Total ✓ $\mathcal{W}$	19.59	23.71	0.00	15.11	21.89	20.45	14.14	15.05	129.94

Table 9. User 3 recording quality and duration.

	Rec	Still	Walk	Run	Bike	Car	Bus	Train	Subway	Tot
✓	070617	0.09	0.19	0.00	0.00	0.00	0.23	0.00	0.00	0.50
✓	080617	1.76	1.00	0.00	0.00	0.00	1.65	1.48	0.00	5.88
✓	m090617	0.00	0.16	0.00	0.00	0.00	0.34	0.00	0.00	0.50
✓$\mathcal{L}$	120617	0.94	1.16	0.27	0.00	0.00	4.14	0.00	0.00	6.51
	m130617	0.06	0.17	0.00	0.00	0.00	0.31	0.00	0.00	0.53
✓$\mathcal{L}$	130617	0.88	1.36	0.00	0.83	0.00	2.85	0.00	0.00	5.93
✓$\mathcal{L}$	140617	1.49	0.90	0.00	1.07	0.00	0.00	1.70	0.00	5.15
✓$\mathcal{L}$	m150617	0.15	0.20	0.00	0.00	0.00	0.27	0.00	0.00	0.62
✓	160617	1.18	0.66	0.00	1.79	0.00	0.00	2.25	0.00	5.88
✓	190617	1.48	1.58	0.27	0.00	0.00	3.14	0.00	0.00	6.46
	m210617	0.00	0.17	0.00	0.00	0.00	0.29	0.00	0.00	0.46
✓	220617	1.17	1.50	0.50	0.00	0.00	3.10	0.00	0.00	6.28
✓	230617	1.72	0.81	0.00	2.28	0.00	0.00	1.70	0.00	6.50
✓$\mathcal{L}$	260617	1.38	1.80	0.25	1.10	0.00	2.38	0.00	0.00	6.91
✓	270617	0.50	0.32	0.00	0.00	0.00	0.00	2.22	2.91	5.94
	280617	1.95	0.68	0.00	1.90	0.00	0.00	1.82	0.00	6.35
✓	m290617	0.27	0.58	0.50	0.00	0.00	0.23	0.00	0.00	1.58
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	300617	0.80	0.73	0.00	1.22	0.00	3.01	0.00	0.00	5.76
✓$\mathcal{C}$ $\mathcal{W}$	030717	0.97	0.29	0.00	0.00	0.00	0.00	2.12	3.08	6.46
✓$\mathcal{C}$ $\mathcal{W}$	040717	1.75	1.51	0.25	1.00	0.00	2.33	0.00	0.00	6.84
✓$\mathcal{C}$ $\mathcal{W}$	050717	0.53	0.90	0.00	0.00	0.00	0.00	2.17	2.91	6.52
✓$\mathcal{C}$ $\mathcal{W}$	m060717	0.00	0.13	0.00	0.00	0.00	0.29	0.00	0.00	0.42
	070717	2.45	1.95	0.52	0.00	1.40	1.38	0.00	0.00	7.70
✓$\mathcal{C}$ $\mathcal{W}$	100717	0.31	0.28	0.00	0.00	0.00	0.00	2.68	2.90	6.16
✓$\mathcal{L}$ $\mathcal{C}$ $\mathcal{W}$	110717	2.26	1.62	0.50	0.98	0.00	0.00	0.00	0.00	5.35
✓$\mathcal{C}$ $\mathcal{W}$	120717	0.65	0.28	0.00	0.00	0.00	0.00	2.19	3.18	6.29
✓$\mathcal{C}$ $\mathcal{W}$	m130717	0.00	0.13	0.00	0.00	0.00	0.30	0.00	0.00	0.43
✓$\mathcal{C}$ $\mathcal{W}$	140717	1.01	0.36	0.00	0.00	0.00	0.00	2.34	2.36	6.08
✓$\mathcal{C}$ $\mathcal{W}$	170717	1.97	2.26	0.50	1.51	0.00	0.00	0.00	0.00	6.25
✓$\mathcal{C}$ $\mathcal{W}$	180717	1.58	2.24	0.00	1.67	0.00	0.00	0.00	0.00	5.48
✓$\mathcal{C}$ $\mathcal{W}$	190717	0.77	0.26	0.00	0.00	0.00	0.00	2.30	2.87	6.20
✓$\mathcal{C}$ $\mathcal{W}$	200717	2.39	1.28	0.50	1.74	0.00	0.00	0.00	0.00	5.91
	210717	0.52	0.32	0.00	0.00	0.00	0.00	2.25	3.27	6.35
✓$\mathcal{C}$ $\mathcal{W}$	240717	1.72	1.39	0.52	1.92	0.00	0.00	0.00	0.00	5.54
✓$\mathcal{C}$ $\mathcal{W}$	m250717	0.00	0.15	0.00	0.00	0.00	0.23	0.00	0.00	0.38
	Total	34.71	29.30	4.58	18.99	1.40	26.47	27.22	23.48	166.14
	Total ✓	29.73	26.00	4.06	17.09	0.00	24.49	23.15	20.21	144.74
	Total ✓$\mathcal{L}$	7.92	7.76	1.02	5.19	0.00	12.65	1.70	0.00	36.23
	Total ✓$\mathcal{C}$	16.72	13.81	2.26	10.03	0.00	6.17	13.80	17.30	80.09
	Total ✓$\mathcal{W}$	16.72	13.81	2.26	10.03	0.00	6.17	13.80	17.30	80.09

3.3 Data collection notes

The firmware of the phones of User 1 has been updated on 11 May 2017. All the recordings up to and including 100517 (10.05.2017) were performed with a firmware which sampled the motion sensors and pressure sensor at exactly 100Hz. Starting from recording 110517 (11.05.2017), the firmware samples the Accelerometer and Gyroscope at 200Hz (screen off) or 250Hz (screen on) and the pressure sensor at 10Hz; the other motion sensors remain sampled at 100Hz. For ease of use, the release dataset provides the data in a resampled format for all the motion sensors and the pressure sensor at 100Hz.

Recordings of User 1 until 190417 and of User 2 and 3 until 300617 relied on "best effort" for Wifi and cell scan, after which we forced to report Wifi and cell coverage every second. This however does not guarantee that the underlying drivers scan Wifi or cells more frequently.

4. Data organisation and file format

The data is organised in the following directory structure and with the following files:

/root/	Root folder of the dataset
/root/<userid>/	Root of the User userid.
/root/<userid>/qc.pdf	Quality check document
/root/<userid>/datasetstatus.mat	Convenience Matlab file containing basic info about the recordings in this session including start time and duration of each session and the quality statistics used to create qc.pdf.
/root/<userid>/labelstatus.mat	Convenience Matlab file containing basic info about the labels of the recording. This infomation is as well available in the files `labels_track_xxx.txt`.
/root/<userid>/<recordid>/	Root of the record. `<recordid>` is generally the date of the recording. All the data files corresponding to that recording are under that folder
/root/<userid>/<recordid>/<position>_Motion.txt	nsx23 column matrix comprising motion and pressure data, with ns the number of samples. `<position>` is one of "Hand", "Bag", "Hips", "Torso".
/root/<userid>/<recordid>/<position>_API.txt	nsx11 column with Google recognition data
/root/<userid>/<recordid>/<position>_Ambient.txt	nsx6 column with Ambient light data
/root/<userid>/<recordid>/<position>_Battery.txt	nsx5 column with the Battery data
/root/<userid>/<recordid>/<position>_GPS.txt	nsxvar column with the GPS data. The number of columns varies depending on the number of visible satellites.
/root/<userid>/<recordid>/<position>_Wifi.txt	nsxvar column with Wifi data. The number of columns is function of the number of nearby Wifis.
/root/<userid>/<recordid>/<position>_Cells.txt	nsxvar column with cell data. The number of columns is function of the number of nearby cells.
/root/<userid>/<recordid>/<position>_DeprCells.txt	nsx9 column with cell data from the deprecated Android API.
/root/<userid>/<recordid>/<position>_Location.txt	nsx7 column indicating latitude/longitude
/root/<userid>/<recordid>/Label.txt	nsx8 matrix with 7 label tracks. The number of samples ns is identical to that in `<position>_Motion.txt` (each line of `<position>_Motion.txt` corresponds to the same line in `Label.txt`).
/root/<userid>/<recordid>/GPS.kml	Google Earth GPS trace

/root/<userid>/<recordid>/timelapse.avi	Body-camera timelapse video.
/root/<userid>/<recordid>/labels_track_main.txt	Main label track
/root/<userid>/<recordid>/labels_track_road.txt	Road label track
/root/<userid>/<recordid>/labels_track_social.txt	Social label track
/root/<userid>/<recordid>/labels_track_tunnels.txt	Tunnels label track
/root/<userid>/<recordid>/labels_track_traffic.txt	Traffic label track
/root/<userid>/<recordid>/labels_track_food.txt	Food label track
/root/<userid>/<recordid>/data.mat	Convenience Matlab file containing a data structure with the pre-parsed data of the variable-width files and the data text files of small size.
/root/<userid>/<recordid>/datainf.mat	Convenience Matlab file containing a data structure with basic info about the recording.

The data of all the motion sensors and the pressure sensors are all combined in a single file `<position>_Motion.txt`. This file is resampled on a regular sampling grid which is identical for all the `<position>` and for `Label.txt`. This simplifies sliding window data processing on these regularly sampled data.

4.1 File formats

4.1.1 <position>_Motion.txt

This file contains one line per sample, resampled on a regular sampling grid with identical sample times for all positions. Some columns may contain NaN if the information is not available (e.g. not all sensors start sampling at the exact same time). The columns are as follows:

Column	What
1	Time ms
2	Acceleration X
3	Acceleration Y
4	Acceleration Z
5	Gyroscope X
6	Gyroscope Y
7	Gyroscope Z
8	Magnetometer X
9	Magnetometer Y
10	Magnetometer Z
11	Orientation q0
12	Orientation q1
13	Orientation q2
14	Orientation q3
15	Gravity X
16	Gravity Y
17	Gravity Z
18	Linear acceleration X
19	Linear acceleration Y
20	Linear acceleration Z
21	Pressure mbar
22	Pressure altitude (0 from recording 110517 of User1 due to new firmware not providing this data)
23	Pressure temperature (0 from recording 110517 of User1 due to new firmware not providing this data)

4.1.2 <position>_API.txt

This file contains one line per sample. The columns are as follows:

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Still confidence (0-100)
5	On foot confidence (0-100)
6	Walking confidence (0-100)
7	Running confidence (0-100)
8	On bicycle confidence (0-100)
9	In vehicle confidence (0-100)
10	Tilting confidence (0-100)
11	Unknown confidence (0-100)

4.1.3 <position>_Ambient.txt

This file contains one line per sample. The columns are as follows:

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Lumix
5	Temperature
6	Ignore

4.1.4 <position>_Battery.txt

This file contains one line per sample. The columns are as follows:

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Battery level
5	Temperature

4.1.5 <position>_GPS.txt

This file contains one line per sample. The columns are as follows:

Column	What
1	Time ms
2	Ignore
3	Ignore

4+	Variable number of entries for GPS data. If no satellite is visible the 4th column is 0.
	Otherwise, for each satellite visible 4 columns are added to the data file and an additional last column indicates the number of satellites.
	Each of the 4 columns contain in order: ID, SNR, Azimuth, Elevation
	For example:
	`1489485950011 161777247369 10889909374 0`
	indicates no satellite visible.
	`1489485951014 162780045286 10889909374 7 12.0 56.0 32.0 1`
	indicates one satellite visible; satellite 7 with SNR=12, Azimuth=56 and elevation=32.
	`1489485962025 173791715076 10889909374 7 15.0 56.0 32.0 30 12.0 82.0 70.0 2`
	indicates two satellite visible; satellite 7 and 30.

4.1.6 <position>_Wifi.txt

This file contains one line per sample. The columns are semicolon delimitated as Wifi SSID can contain spaces. The columns are as follows

Column	What
1	Time ms
2	Ignore
3	Ignore
4+	Variable number of Wifi data. For each visible Wifi 5 semicolon delimited fields are included, in order: BSSID, SSID, RSSI, Frequency, Capabilities.

4.1.7 <position>_DeprCells.txt

This file contains one line per sample. Some columns may contain NaN if the information is not available (e.g. the signal is too weak or the cell is not registered). The columns are as follows:

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Network type
5	cid
6	lac
7	dBm
8	MCC
9	MNS

4.1.8 <position>_Cells.txt

This file contains one line per sample. The columns are as follows

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Number of entries
5+	Variable number of fields depending on entries. The first field identifies the type of cell: LTE, WCDMA or GSM. The number of fields depends on the cell type and is as follows: <LTE; Signal level; Signal strength; Signal level; 28-bit Cell Identity; 3-digit Mobile Country Code; 2 or 3-digit Mobile Network Code; Physical Cell Id; 16-bit Tracking Area Code> <GSM; Signal level calculated based on 3GPP RSRP; Signal strength; Signal level; 16-bit GSM Cell Identity described in TS 27.007; 16-bit Location Area Code; 3-digit Mobile Country Code; 2 or 3-digit Mobile Network Code> <WCDMA; isRegistered; cid; lac; MCC; MNC; PSC; asuLevel; dBm; level>

4.1.9 <position>_Location.txt

This file contains one line per sample. The columns are as follows

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Accuracy of this location (accuracy as the radius of 68% confidence)
5	Latitude
6	Longitude
7	Altitude

4.1.10 Label.txt

This file contains one line per sample. This file is derived from `labels_track_xxx.txt` (with xxx being main, road, social, tunnels, traffic, food) and is provided as a convenience. Each line of the file corresponds to the same line in `<position>_Motion.txt`. This may simplify some streaming processing. The column 1 in `Label.txt` is identical to column 1 in `<position>_Motion.txt`.

Column	What	
1	Time [ms]	
2	Main label:	Null=0, Still=1, Walking=2, Run=3, Bike=4, Car=5, Bus=6, Train=7, Subway=8
3	Fine label:	
	Null=	0
	Still;Stand;Outside=	1
	Still;Stand;Inside=	2
	Still;Sit;Outside=	3
	Still;Sit;Inside=	4
	Walking;Outside=	5
	Walking;Inside=	6
	Run=	7
	Bike=	8
	Car;Driver=	9
	Car;Passenger=	10
	Bus;Stand=	11
	Bus;Sit=	12
	Bus;Up;Stand=	13
	Bus;Up;Sit=	14
	Train;Stand=	15
	Train;Sit=	16
	Subway;Stand=	17
	Subway;Sit=	18
4	Road label:	City=1, Motorway=2, Countryside=3, Dirt road=4, Null=0
5	Traffic label:	Heavy traffic=1, null=0
6	Tunnels label:	Tunnel=1, null=0
7	Social label:	Social=1, null=0
8	Food label:	Eating=1, Drinking=2, Both=3, null=4

4.1.11 labels_track_main.txt

This file contains one line per label. The columns are as follows

Column	What	
1	Label start time in millisecond	
2	Label end time in millisecond	
3	Activity label	
	Still;Stand;Outside:	0
	Still;Stand;Inside:	1
	Still;Sit;Outside:	2
	Still;Sit;Inside:	3
	Walking;Outside:	4
	Walking;Inside:	5
	Run:	6
	Bike:	7
	Car;Driver:	8
	Car;Passenger:	9
	Bus;Stand:	10
	Bus;Sit:	11
	Bus;Up;Stand:	12
	Bus;Up;Sit:	13
	Train;Stand:	14
	Train;Sit:	15
	Subway;Stand:	16
	Subway;Sit:	17

4.1.12 labels_track_road.txt

This file contains one line per label. The columns are as follows

Column	What
1	Label start time in millisecond
2	Label end time in millisecond
3	Label: City: 0 Motorway: 1 Countryside: 2 Dirt road: 3

4.1.13 labels_track_traffic.txt, labels_track_tunnels.txt, labels_track_social.txt

This file contains one line per label. The columns are as follows

Column	What
1	Label start time in millisecond
2	Label end time in millisecond
3	Label: 1=yes (heavy traffic, in tunnel, social interactions) 0=no

4.1.14 labels_track_food.txt

This file contains one line per label. The columns are as follows

Column	What
1	Label start time in millisecond
2	Label end time in millisecond
3	Label: Eating: 0 Drinking: 1 Both: 2

4.1.15 data.mat

Convenience matlab file containing pre-parsed data and data of text files of small size. Loading data.mat yields a data variable which is a structure comprising the fields described in the following table.

Each field comprises a 4-entry cell corresponding to each of the 4 body-location of the phones. Locations 1 to 4 correspond to Hand, Bag, Hips and Torso respectively.

Fields	What
Battery{p}	nx5 array comprising the battery data for position p. Format identical to <p>_Battery.txt
Ambient{p}	nx5 array comprising the Ambient data for position p. Format identical to <p>_Ambient.txt
API{p}	nx11 array comprising the recognition API data for position p. Format identical to <p>_Ambient.txt
Location{p}	nx7 array comprising the Location data for position p. Format identical to <p>_Location.txt
Cells{p}	nx7 array comprising the following columns: 1: Time [ms] 2: Ignore 3: Ignore 4 Total number of cells 5 number of LTE cells 6 Number of GSM cells 7 Number of WCDMA
Cellsvar{p}	n-entry cell containing the textual description of the cells. Each entry Cellsvar{p}{i} corresponds to the same entry in Cells{p}{i}. This corresponds to the variable length part of <position>_Cells.txt.
Wifi{p}	nx4 array comprising the following columns: 1 Time [ms] 2 Ignore 3 Ignore 4 Number of Wifi access points

Wifivar{p}	n-entry cell containing the textual description of the Wifi access points. Each entry Wifivar{p}{i} corresponds to the same entry in Wifi{p}{i}. This corresponds to the variable length part of <position>_Wifi.txt.
GPS{p}	nxvar array with GPS data. The columns are as follows: 1 Time [ms] 2 Ignore 3 Ignore 4 Number N of GPS satellites 5-8 Data for 1st GPS: ID, SNR, Azimuth, Elevation 9-12 Data for 2nd GPS, etc.
sessionid	Name of the session (User1, User2 or User3)
timemsmin	Time of start of recording in milliseconds.
timemsmax	Time of end of recording in milliseconds.
recdate	String with date/time of start of recording
reclength	Duration of recording in milliseconds
recid	Recording ID, e.g. '010317'.

4.1.16 datainf.mat

Convenience matlab file containing basic info about the recording. Loading datainf.mat yields a datainfo variable which is a structure comprising the fields described in the following table.

Fields	What
sessionid	Name of the session (User1, User2 or User3)
timemsmin	Time of start of recording in milliseconds.
timemsmax	Time of end of recording in milliseconds.
recdate	String with date/time of start of recording
reclength	Duration of recording in milliseconds
recid	Recording ID, e.g. '010317'.

4.1.17 datasetstatus.mat

Convenience matlab file containing basic info about all the recordings of the user. Loading datasetstatus.mat yields a datasetstatus variable which is a map. The key to the map is the recording ig (e.g. '010317') and the value is a structure comprising the fields described in the following table:

Fields	What
sessionid	Name of the session (User1, User2 or User3)
timemsmin	Time of start of recording in milliseconds.
timemsmax	Time of end of recording in milliseconds.
recdate	String with date/time of start of recording
reclength	Duration of recording in milliseconds
recid	Recording ID, e.g. '010317'.
status	Structure comprising status information as follows:

Fields	What
dtok	Structure indicating if the time interval between samples appears OK according to the automatic checks. This structure contains fields for each sensors (e.g. Accelerometer, Gyroscope). Each field is an 4x1 matrix comprising the status for each of the four phone positions, with index 1 to 4 correspond to Hand, Bag, Hips and Torso respectively.
covok	Structure indicating if the time sensor coverage appears OK according to the automatic checks.
coverage	Structure indicating the coverage of the sensor with 1 indicating that the number of samples is as expected and values higher or lower than 1 indicating more or less samples than expected.
boundaryok	Structure indicating whether the first and last sample of a sensor align with the start and the end of the recording within a tolerance window.
ok	Structure indicating whether the sensor data is ok; this is a combination of dtok, covok and boundaryok.
restarts	4x1 matrix indicating the number of restarts of each phones.
motion200hz	0 if the motion sensors are sampled at 100Hz or 1 if the motion sensors are sampled at 200Hz/250Hz.

4.1.18 labelstatus.mat

Convenience matlab file containing basic info about all the recordings of the user. Loading labelstatus.mat yields a labelstatus variable which is a map. The key to the map is the recording ig (e.g. '010317') and the value is a structure comprising the fields described in the following table:

Fields	What
main	nx3 array. This is the data of labels_track_main.txt.
food	nx3 array. This is the data of labels_track_food.txt.
road	nx3 array. This is the data of labels_track_road.txt.
tunnels	nx3 array. This is the data of labels_track_tunnels.txt.
social	nx3 array. This is the data of labels_track_social.txt.
traffic	nx3 array. This is the data of labels_track_traffic.txt.
dstill	Duration of still in the recording in milliseconds
dwalk	Duration of walk in the recording in milliseconds
drun	Duration of run in the recording in milliseconds
dbike	Duration of bike in the recording in milliseconds
dcar	Duration of car in the recording in milliseconds
dbus	Duration of bus in the recording in milliseconds
dtrain	Duration of train in the recording in milliseconds
dsubway	Duration of subway in the recording in milliseconds
dtow	Total duration of annotations in the recording in milliseconds.
sessionid	Name of the session (User1, User2 or User3)
recid	Recording ID, e.g. '010317'.