The ExtraSensory Dataset

A dataset for behavioral context recognition in-the-wild from mobile sensors

Jump to: Dataset description How was the data collected? Download Tutorial Open problems Relevant papers

- Large scale: over 300k examples (minutes) from 60 users.
- Everyday devices: sensors from smartphone (iPhone/Android) and smartwatch.
- Diverse sensors: heterogeneous measurements from different sensors.
- In-the-Wild: data was collected from users that were engaged in their regular natural behavior.
- Rich context: annotations are combinations of context labels from a large vocabulary.
- Publicly available: everyone is invited to download the dataset for free and use it (conditioned on citing our original paper).

This dataset was collected in 2015-2016 by Yonatan Vaizman and Katherine Ellis with the supervision of professor Gert Lanckriet.
Department of Electrical and Computer Engineering, University of California, San Diego.
Original publication: "Recognizing Detailed Human Context In-the-Wild from Smartphones and Smartwatches".

Dataset description:

The ExtraSensory dataset contains data from 60 users (also referred to as subjects or participants), each identified with a universally unique identifier (UUID). From every user it has thousands of examples, typically taken in intervals of 1 minute (but not necessarily in one long sequence, there are time gaps). Every example contains measurements from sensors (from the user's personal smartphone and from a smartwatch that we provided). Most examples also have context labels self-reported by the user.

Users:
The users were mostly students (both undergraduate and graduate) and research assistants from the UCSD campus.
34 iPhone users, 26 Android users.
34 female, 26 male.

56 right handed, 2 left handed, 2 defined themselves as using both.
Diverse ethnic backgrounds (each user defined their "ethnicity" how they liked), including Indian, Chinese, Mexican, Caucasian, Filipino, African American and more.
Here are some more statistics over the 60 users:

	Range	Average (standard deviation)
Age (years)	18-42	24.7 (5.6)
Height (cm)	145-188	171 (9)
Weight (kg)	50-93	66 (11)
Body mass index (kg/m^2)	18-32	23 (3)
Labeled examples	685-9,706	5,139 (2,332)
Additional unlabeled examples	2-6,218	1,150 (1,246)
Average applied labels per example	1.1-9.7	3.8 (1.4)
Days of participation	2.9-28.1	7.6 (3.2)

Devices:
The users in ExtraSensory had a variety of phone devices.
iPhone generations: 4, 4S, 5, 5S, 5C, 6 and 6S.
iPhone operating system versions ranging from iOS-7 to iOS-9.
Android devices: Samsung, Nexus, HTC, moto G, LG, Motorola, One Plus One, Sony.

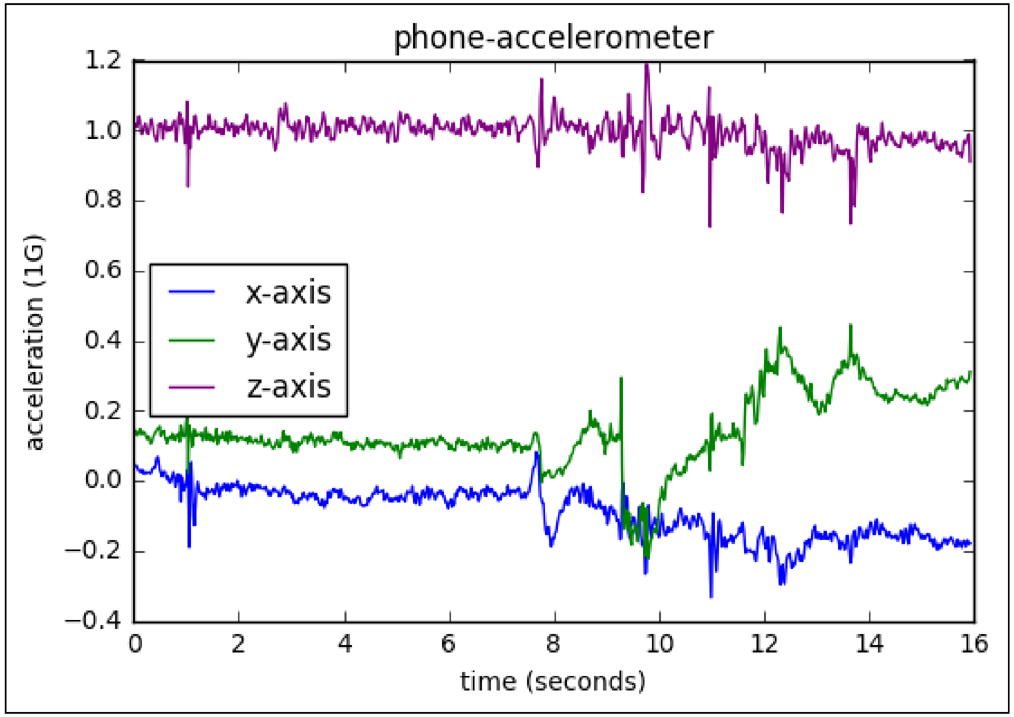
Sensors:
The sensors used were diverse and include high-frequency motion-reactive sensors (accelerometer, gyroscope, magnetometer, watch accelerometer), location services, audio, watch compass, phone state indicators and additional sensors that were sampled in low frequency (once a minute).
Not all sensors were available all the time. Some phones didn't have some sensors (e.g. iPhones didn't have air pressure sensor). In other cases sensors were sometimes unavailable (e.g. location services were sometimes turned off by the user's choice, audio was not available when the user was on a phone call).
The following table specifies the different sensors, the format of their measurements for a single example and the total number of labeled examples (#ex) and users (#us) that have measurements from each sensor.

sensor	details	dimension	#us	#ex
accelerometer	Tri-axial direction and magnitude of acceleration. 40Hz for ~20sec.	(~800) x 3	60	308,306
gyroscope	Rate of rotation around phone's 3 axes. 40Hz for ~20sec.	(~800) x 3	57	291,883
magnetometer	Tri-axial direction and magnitude of magnetic field. 40Hz for ~20sec.	(~800) x 3	58	282,527
watch accelerometer	Tri-axial acceleration from the watch. 25Hz for ~20sec.	(~500) x 3	56	210,716
watch compass	Watch heading (degrees). nC samples (whenever changes in 1deg).	nC x 1	53	126,781
location	Latitude, longitude, altitude, speed, accuracies. nL samples (whenever changed enough).	nL x 6	58	273,737

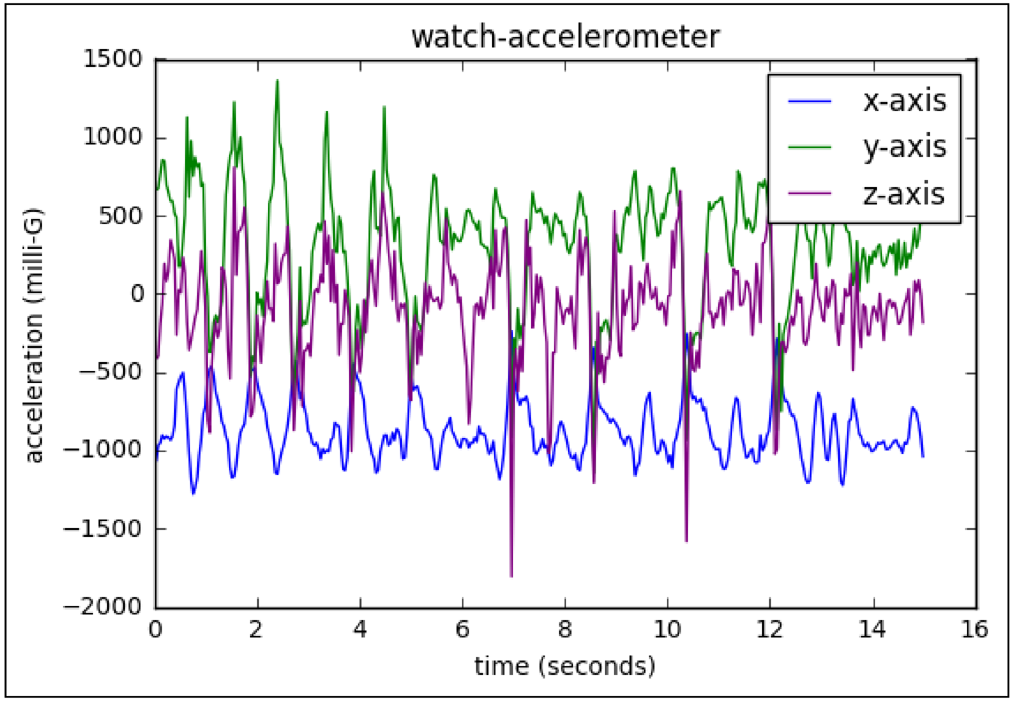
location (quick)	Quick location-variability features (no absolute coordinates) calculated on the phone.	1 x 6	58	263,899
audio	22kHz for ~20sec. Then 13 MFCC features from half overlapping 96msec frames.	(~430) x 13	60	302,177
audio magnitude	Max absolute value of recorded audio, before it was normalized.	1	60	308,877
phone state	App status, battery state, WiFi availability, on the phone, time-of-day.	5 discrete	60	308,320
additional	Light, air pressure, humidity, temperature, proximity. If available sampled once in session.	5	---	---

Here are some examples of raw-measurements recorded from various sensors during the 20-second window. These examples are taken from different examples in the dataset (the relevant context is presented in parenthesis):

Phone-accelerometer (recorded while running with phone in pocket):

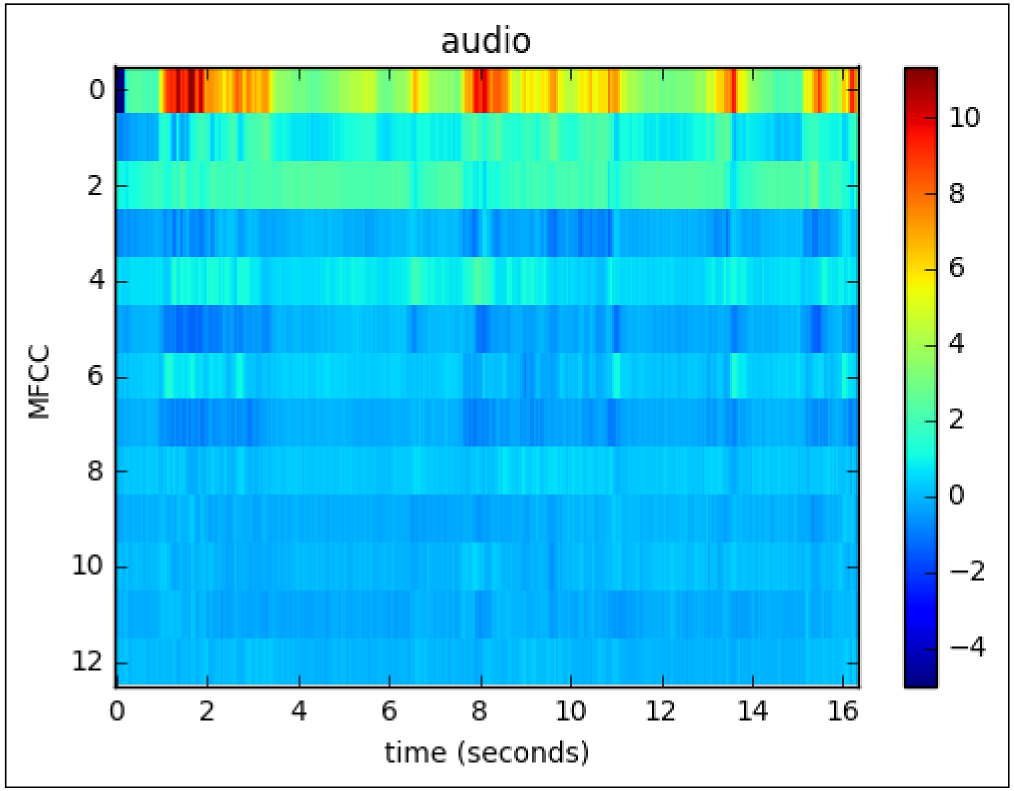


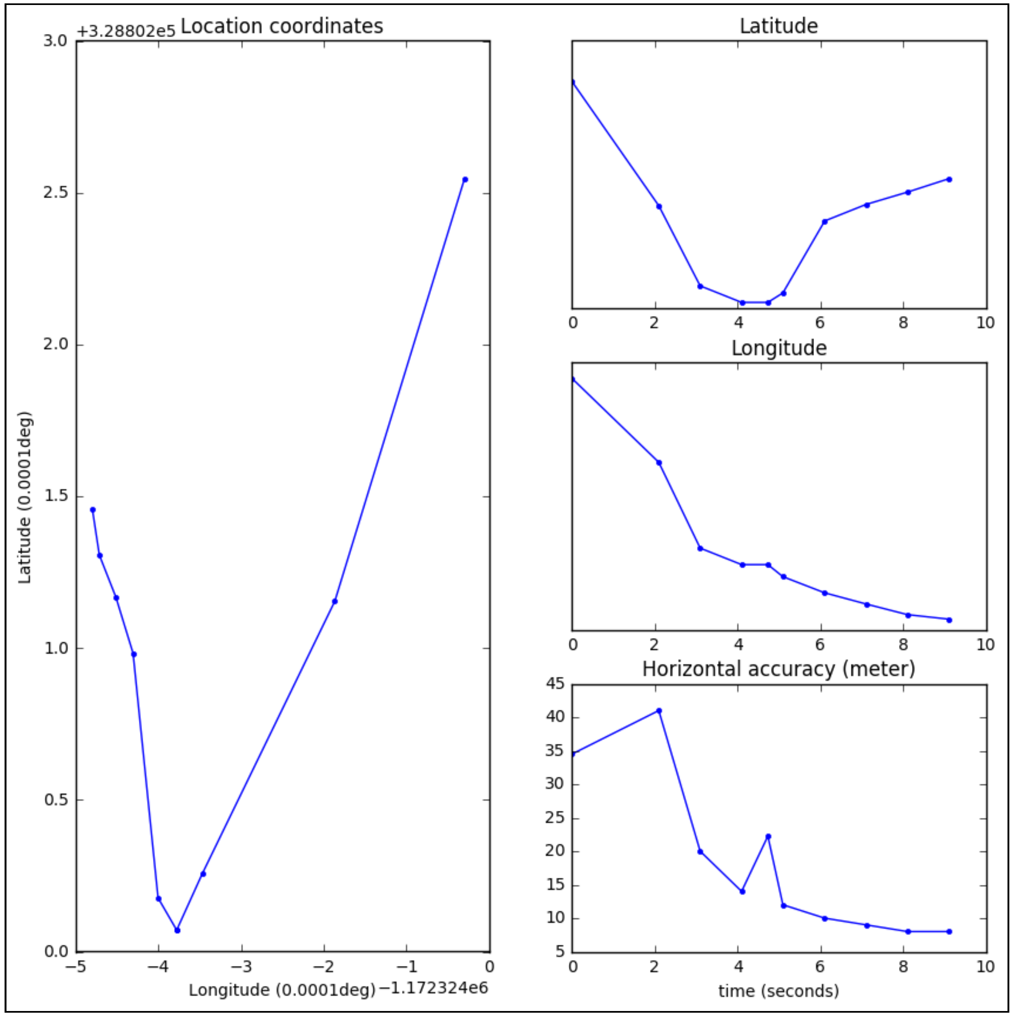
Watch-accelerometer (recorded during shower):



Audio (recorded while watching TV and eating at home):

Location (recorded during drive in a car):





Additional measurements were recorded from pseudo-sensors - processed versions that are given by the OS:

- Calibrated version of gyroscope (tries to remove drift affect).
- Unbiased version of magnetometer (tries to remove bias of the magnetic field created by the phone itself).
- Gravitation direction (the magnitude is always 1G).
- User-generated acceleration (raw acceleration minus gravitation acceleration).
- Estimated orientation of the phone.
- Rotation vector or attitude vector at every time-point.

Labels:
Cleaned labels:

We have processed and cleaned the labels that were self-reported by users.
Labels with prefix 'OR_', 'LOC_', or 'FIX_' are processed versions of original labels.
The primary data provided here is from these cleaned labels, including the following (sorted according to descending order of number of examples):

	Label	#users	#examples
1	OR_indoors	59	184692
2	LOC_home	57	152892
3	SITTING	60	136356
4	PHONE_ON_TABLE	53	115037
5	LYING_DOWN	58	104210
6	SLEEPING	53	83055
7	AT_SCHOOL	49	42331
8	COMPUTER_WORK	45	38081
9	OR_standing	60	37782
10	TALKING	54	36293
11	LOC_main_workplace	32	33944
12	WITH_FRIENDS	32	24737
13	PHONE_IN_POCKET	40	23401
14	FIX_walking	60	22136
15	SURFING_THE_INTERNET	35	19416
16	EATING	57	16594
17	PHONE_IN_HAND	43	14573
18	WATCHING_TV	40	13311
19	OR_outside	45	12114
20	PHONE_IN_BAG	26	10201
21	OR_exercise	44	8081
22	DRIVE_-_I_M_THE_DRIVER	31	7975
23	WITH_CO-WORKERS	21	6224
24	IN_CLASS	20	6110
25	IN_A_CAR	33	6083
26	IN_A_MEETING	45	5153

27	BICYCLING	25	5020
28	COOKING	40	4029
29	LAB_WORK	9	3848
30	CLEANING	30	3806
31	GROOMING	36	3064
32	TOILET	43	2655
33	DRIVE_-_I_M_A_PASSENGER	23	2526
34	DRESSING	35	2233
35	FIX_restaurant	28	2098
36	BATHING_-_SHOWER	37	2087
37	SHOPPING	27	1841
38	ON_A_BUS	31	1794
39	AT_A_PARTY	9	1470
40	DRINKING__ALCOHOL_	12	1456
41	WASHING_DISHES	25	1228
42	AT_THE_GYM	8	1151
43	FIX_running	26	1090
44	STROLLING	11	806
45	STAIRS_-_GOING_UP	18	798
46	STAIRS_-_GOING_DOWN	19	774
47	SINGING	8	651
48	LOC_beach	8	585
49	DOING_LAUNDRY	15	556
50	AT_A_BAR	5	551
51	ELEVATOR	12	200

To better understand how the data is structured, and how to use it, check out our introduction tutorial.

Original labels:
In the mobile app self-reporting interface, the users could report labels of two types:

1. Main activity. Labels describing movement or posture of the user. This category is mutually exclusive and the possible 7 values are: lying down, sitting, standing in place, standing and moving, walking, running, bicycling.
2. Secondary activity. Additional 109 labels describing more specific context in different aspects: sports (e.g. playing basketball, at the gym), transportation (e.g. drive - I'm the driver, on the bus), basic needs (e.g. sleeping, eating, toilet), company (e.g. with family, with co-workers), location (e.g. at home, at work, outside) etc.
   Multiple secondary labels can be applied to an example.

Some examples may have no main activity selected, but have secondary labels (e.g. when the user didn't remember if they were sitting or walking, but did remember they were indoors).

On average (over the sixty users) an example has more than 3 labels assigned to it.
On average a user's label usage distribution has an entropy of 3.9 bits, which roughly mean that a typical user mainly used ~15 labels during the participation period.
The following table displays the labels (main and secondary) and specifies for each label the number of examples that have the label applied and the number of users that used the label. Labels are numbered according to descending order of number of examples.

	Label	**#users**	**#examples**
1	SITTING	60	136356
2	PHONE_ON_TABLE	53	116425
3	LYING_DOWN	58	104210
4	AT_HOME	55	103889
5	SLEEPING	53	83055
6	INDOORS	31	57021
7	AT_SCHOOL	49	42331
8	COMPUTER_WORK	45	38081
9	TALKING	54	36293
10	STANDING_AND_MOVING	58	29754
11	AT_WORK	32	29574
12	STUDYING	33	26277
13	WITH_FRIENDS	32	24737
14	PHONE_IN_POCKET	40	24226
15	WALKING	60	22517
16	RELAXING	32	21223
17	SURFING_THE_INTERNET	35	19416

18	PHONE_AWAY_FROM_ME	27	17937
19	EATING	57	16594
20	PHONE_IN_HAND	43	16308
21	WATCHING_TV	40	13311
22	OUTSIDE	40	11967
23	PHONE_IN_BAG	26	10760
24	LISTENING_TO_MUSIC__WITH_EARPHONES_	31	10228
25	WRITTEN_WORK	15	9083
26	STANDING_IN_PLACE	59	8028
27	DRIVE_-_I_M_THE_DRIVER	31	7975
28	WITH_FAMILY	14	7749
29	WITH_CO-WORKERS	21	6224
30	IN_CLASS	20	6110
31	IN_A_CAR	33	6083
32	TEXTING	24	5936
33	LISTENING_TO_MUSIC__NO_EARPHONES_	24	5589
34	DRINKING__NON-ALCOHOL_	30	5544
35	IN_A_MEETING	45	5153
36	WITH_A_PET	1	5125
37	BICYCLING	25	5020
38	LISTENING_TO_AUDIO__NO_EARPHONES_	11	4359
39	READING_A_BOOK	22	4223
40	COOKING	40	4029
41	LISTENING_TO_AUDIO__WITH_EARPHONES_	7	4029
42	LAB_WORK	9	3848
43	CLEANING	30	3806
44	GROOMING	36	3064
45	EXERCISING	14	2679
46	TOILET	43	2655
47	DRIVE_-_I_M_A_PASSENGER	23	2526

48	AT_A_RESTAURANT	29	2519
49	PLAYING_VIDEOGAMES	9	2441
50	LAUGHING	8	2428
51	DRESSING	35	2233
52	BATHING_-_SHOWER	37	2087
53	SHOPPING	27	1841
54	ON_A_BUS	31	1794
55	STRETCHING	12	1667
56	AT_A_PARTY	9	1470
57	DRINKING__ALCOHOL_	12	1456
58	RUNNING	28	1335
59	WASHING_DISHES	25	1228
60	SMOKING	2	1183
61	AT_THE_GYM	8	1151
62	ON_A_DATE	6	1086
63	STROLLING	11	806
64	STAIRS_-_GOING_UP	18	798
65	STAIRS_-_GOING_DOWN	19	774
66	SINGING	8	651
67	ON_A_PLANE	4	630
68	DOING_LAUNDRY	15	556
69	AT_A_BAR	5	551
70	AT_A_CONCERT	5	538
71	MANUAL_LABOR	8	494
72	PLAYING_PHONE-GAMES	4	403
73	ON_A_TRAIN	5	344
74	DRAWING	3	273
75	ELLIPTICAL_MACHINE	2	233
76	AT_THE_BEACH	6	230

77	AT_THE_POOL	5	216
78	ELEVATOR	12	200
79	TREADMILL	2	164
80	PLAYING_BASEBALL	2	163
81	LIFTING_WEIGHTS	1	162
82	SKATEBOARDING	3	131
83	YOGA	3	128
84	BATHING_-_BATH	6	121
85	DANCING	3	115
86	PLAYING_MUSICAL_INSTRUMENT	2	114
87	STATIONARY_BIKE	2	86
88	MOTORBIKE	1	86
89	TRANSFER_-_BED_TO_STAND	4	73
90	VACUUMING	1	68
91	TRANSFER_-_STAND_TO_BED	4	63
92	LIMPING	1	62
93	PLAYING_FRISBEE	2	54
94	AT_A_SPORTS_EVENT	2	52
95	PHONE_-_SOMEONE_ELSE_USING_IT	3	41
96	JUMPING	1	29
97	PHONE_STRAPPED	1	27
98	GARDENING	1	21
99	RAKING_LEAVES	1	21
100	AT_SEA	1	18
101	ON_A_BOAT	1	18
102	WHEELCHAIR	1	9
103	WHISTLING	1	5
104	PLAYING_BASKETBALL	0	0
105	PLAYING_LACROSSE	0	0
106	PLAYING_SOCCER	0	0

107	MOWING_THE_LAWN	0	0
108	WASHING_CAR	0	0
109	HIKING	0	0
110	CRYING	0	0
111	USING_CRUTCHES	0	0
112	RIDING_AN_ANIMAL	0	0
113	TRANSFER_-_BED_TO_WHEELCHAIR	0	0
114	TRANSFER_-_WHEELCHAIR_TO_BED	0	0
115	WITH_KIDS	0	0
116	TAKING_CARE_OF_KIDS	0	0

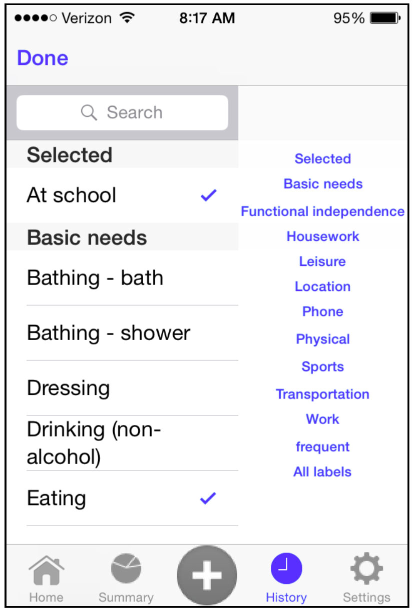
Out of the main and secondary labels 103 were applied by users. Although some labels were applied very rarely they may still be useful, for instance by joining labels with logical-or operation (e.g. "running or playing Frisbee or playing baseball"). In addition there are cases where the user wrongfully applied an irrelevant label; and more commonly, cases where a relevant label wasn't reported by the user (e.g. at home). For those reasons we conducted the cleaning and provide the cleaned-version of the labels.

How was the data collected?

Data was collected using the ExtraSensory mobile application. We developed a version for iPhone and a version for Android, with a Pebble watch component that interfaces with both the iPhone and the Android versions. The app performs a 20-second "recording session" automatically every minute. In every recording session the app collects measurements from the phone's sensors and from the watch (if it is available), including: the phone's accelerometer, gyroscope and magnetometer (sampled in 40Hz), audio (sampled in 22kHz, then processed to MFCC feature representation), location, the watch's accelerometer (sampled in 25Hz) and compass and additional sensors if available (light, humidity, air pressure, temperature). The measurements from a recording session are bundled into a zip file and sent to the lab's web server (if WiFi is available, or stored on the phone until WiFi is available).

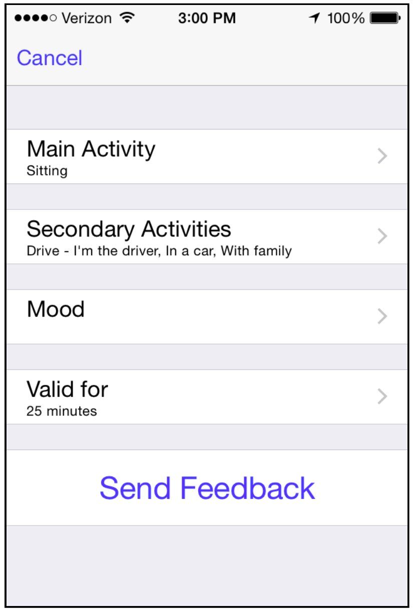
In addition, the app's interface is flexible and has many mechanisms to allow the user to report labels describing their activity and context:

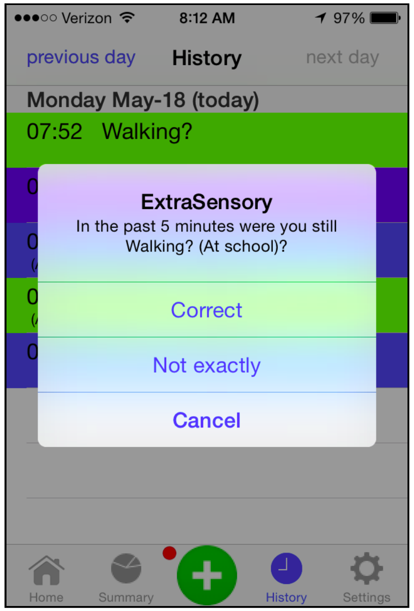
History view. Designed as a daily journal where every item is a chapter of time where the context was the same. Real-time predictions from the server are sent back to the phone and appear on the history as a basic "guess" of the main activity (with question mark). By clicking on an item in the history the user can provide their actual context labels by selecting from menus (including multiple relevant labels together). The user can also easily merge consecutive history items to a longer time-period when the context was constant, or split an item in case the context changed during its period of time. The user can view the daily journal of previous days, but can only edit labels for today and yesterday.



Label selection view. The interface to select the so called 'secondary activity' has a large menu of over 100 context labels. The user has the option to select multiple labels simultaneously. In order to easily and quickly find the relevant labels, the menu is equipped with a side-bar index. The user can find a relevant label in the appropriate index-topic(s), e.g. 'Skateboarding' can be found under 'Sports' and under 'Transportation'. The 'frequent' index-topic is a convenient link to show the user their own personalized list of frequently-used labels.

Active feedback view. The user can report the relevant context labels for the immediate future and report that the same labels will stay relevant for a selected amount of time (up to 30 minutes).





Notifications. Periodically (every 20 minutes by default, but the user can set the interval) the app raises a notification to the user. In case no labels were reported in a while, the notification will ask the user to provide labels. In case the user reported labels recently, the notification will ask whether the context remained the same until now. The notification also appear on the face of the smartwatch, and if the context labels remain the same a simple click of a watch-button is sufficient to apply the same labels for all the recent minutes.

The labels are also sent to the lab's server and saved with the sensor-measurements.

We conducted a meeting with every participant, in which we installed the app on their *personal* phone and explained how to use the app. We provided the Pebble watch to the participant for the week of study, as well as an external battery to allow them for an extra charge of the phone during the day (because the app takes much of the battery). We requested the participant to engage in their regular natural behavior while the app is recording and to try to report as many labels as they can without it bothering their natural behavior too much.

For full details on how we collected the dataset, please refer to our original paper, "Recognizing Detailed Human Context In-the-Wild from Smartphones and Smartwatches".

Download the dataset:

If you use the dataset for your published work, you are required to cite the ExtraSensory original publication, mentioned here as Vaizman2017a.

1. **Primary data - features and labels**.
   The zip file contains a separate 'csv.gz' file for each user in the dataset.
   Each user's csv file (after uncompressing the gzip format) holds all the examples for that user.
   From each example, there are features computed from the different sensors, and the cleaned context labels.
   Download the features and labels zip file (215MB).
   README for the features and labels data.
   Read the tutorial to better understand how to use the data in these files.
2. **Cross validation partition**.
   Download this if you want to perform classification experiments and evaluate them.
   This has a pre-generated partition of the 60 users in the data to 5-folds and prepared text files with the list of users (UUIDs) of the train set and test set of each fold.
   This is the same partition that was used for the experiments in the original paper, Vaizman2017a.
   Use this also to see all the UUIDs and see which users used iPhone and which used Android.
   Download the cross-validation partition zip file.

   Additional parts of the data:

3. **Original context labels**.
   The original labels, as were self-reported by the users.
   This version has the full label list from the mobile app interface. The labels here have only two values: either "reported" or "not reported" (they don't have the notion of "missing labels" that the cleaned labels have).
   This version of the labels is less reliable (it's before cleaning), but you can still use it. It also includes additional interesting labels that were not included in the cleaned labels, like "LISTENING_TO_MUSIC__NO_EARPHONES_" or "PLAYING_VIDEOGAMES".
   The zip file contains a separate 'csv.gz' file for each user in the dataset.
   Download the original labels zip file (970KB).
4. **Mood labels**.
   Although the data collection app (ExtraSensory App) had an option to select mood labels when annotating, we did not focus data collection on mood.

We told the users that they do not have to report mood, but they can only if it is really clear to them how they felt.
Only few users reported mood labels, so some of the mood label data files are filled with missing labels.
Download the mood labels zip file (795KB).

5. **Absolute location coordinates**.
   If you are interested in the absolute geographic location (the actual latitude-longitude coordinates) data you can use this part of the data.
   (The primary data has location-features that are only based on relative-location, meaning they only refer to the variability in space during the recorded 20-second window).
   The zip file contains a separate 'csv.gz' file for each user in the dataset, holding the latitude and longitude coordinates for the user's examples (indexed by timestamp).
   The original location measurements that were recorded by the mobile app during every example's 20-second window were in the form of a sequence of location-updates (new update every time there's a significant change in location).
   The coordinates we give here are 'representative' coordinates for an example and they were calculated as followed:
   - Value of 'nan' if the example has no location coordinate measurements, or if all the location updates taken in this example have poor horizontal-accuracy (more than 200meter).
   - From the updates with best (minimal) horizontal-accuracy, take the coordinates of the latest update.

   Download the absolute location zip file (2.2MB).
6. **Raw sensor measurements**.
   If you are interested in the signal processing stages to extract features, or feature learning, you may want to work with the raw sensor measurements.
   The following files are separated by sensor and inside each zip file the measurment files are arranged by user (UUID) and timestamp.
   Notice that not all sensors were available for recording at all times, so for some examples (timestamps) the measurement file may be missing or may be a dummy file containing just 'nan'.
   Be gentle on the server! The raw-measurement files are very large, so please only download one at a time.
   - Accelerometer measurements: Download the phone-accelerometer measurements zip file (6.1GB).
   - Gyroscope (calibrated) measurements: Download the calibrated gyroscope measurements zip file (8.7GB).
   - Magnetometer measurements: Download the raw (uncalibrated) magnetometer measurements zip file (3.8GB).
   - Watch-accelerometer measurements: Download the watch-accelerometer measurements zip file (800MB).
   - Watch-compass measurements (heading): Download the watch-compass measurements zip file (88MB).
   - Audio measurements: Download the audio measurements zip file - this is not raw audio signals, but rather MFCCs computed on the phone (11GB).
   - Gravity measurements: Download the measurements of the estimated gravity component of the phone-acceleration. Zip file (9.4GB).

Tutorial:

To help you get familiar and comfortable with the dataset, we provide a tutorial, with Python code, in the form of an i-Python notebook:
Download the tutorial in i-Python notebook.
Alternatively, you can view the output of the notebook as a document, in the following page:
Introduction to ExtraSensory (html version).

View this lecture by Yonatan Vaizman for an introduction to **Behavioral Context Recognition** and specifically to the ExtraSensory App and the ExtraSensory Dataset.

Open problems:

The ExtraSensory Dataset enables research and development of algorithms and comparison of solutions to many problems, related to behavioral context recognition. Here are some of the related problems, some of them were addressed in our papers, others remain open for you to solve:

Sensor fusion	The dataset has features (and raw measurements) from sensors of diverse modalities, from the phone and from the watch. In Vaizman2017a (referenced below), we compared different approaches to fuse information from the different sensors, namely early-fusion (concatenation of features) and late-fusion (averaging or weighted averaging of probability outputs from 6 single-sensor classifiers).
Multi-task modeling	The general context-recognition task in the ExtraSensory Dataset is a multi-label task, where at any minute the behavioral context can be described by a combination of relevant context-labels. In Vaizman2017b (referenced below), we compared the basline system of separate model-per-label with a multi-task MLP that outputs probabilities for 51 labels. We showed the advantage of sharing parameters in a unified model. Specifically, an MLP with narrow hidden layers can be richer than a linear model, while having fewer parameters, thus reducing over-fitting. Perhaps other methods can also successfully model many diverse context-labels in a unified model.
Absolute location	The ExtraSensory Dataset includes location coordinates for many examples. So far, in our papers, we only extracted *relative location* features - capturing how much a person moves around in space within each minute. We did not address utilizing the *absolute location* data. There may be useful information in addressing the movement from minute to minute, and incorporating GIS data and geographic landmarks.
Time series modeling	The models we suggested so far treat each example (minute) as independent of the others. There's a lot of work to be done on modeling minute-by-minute time series, smoothing the recognition over minutes, and ways to segment time into meaningful "behavioral events".
More sensing modalities	The dataset includes occasional measurements from sensors that we did not yet utilize in our experiments, including magnetometer, ambient light, air pressure, humidity, temperature, and watch-compass.
Semi-supervised learning	The ability to improve a model with plenty unlabeled examples will enable collecting huge amounts of data with little effort (less self-reporting).
Active learning	Active learning will make future data collections easier on participants - instead of asking for labels for many examples, the system can sparsely prompt the user for labels in the most critical examples.
User adaptation	In Vaizman2017a (referenced below), we demonstrated the potential improvement of context-recognition with a few days of labeled data from a new users. Can you achieve successful user-adaptation *without labels* from the new user?
Feature learning	All our experiments were done with designed features, using traditional DSP methods.

	Feature learning can potentially extract meaningful information from the sensor measurements that the designed features miss. The dataset includes the full raw measurements from the sensors and enables experimenting with feature learning.
Privacy	The ExtraSensory Dataset can be a testbed to compare methods for privacy-preserving.

Relevant papers:

Vaizman2017a	**Vaizman, Y., Ellis, K., and Lanckriet, G. "Recognizing Detailed Human Context In-the-Wild from Smartphones and Smartwatches". *IEEE Pervasive Computing*, vol. 16, no. 4, October-December 2017, pp. 62-74.** *Cite this paper if you use the dataset for any publication!	Accepted version Supplementary
Vaizman2017b	**Vaizman, Y., Weibel, N., and Lanckriet, G. "Context Recognition In-the-Wild: Unified Model for Multi-Modal Sensors and Multi-Label Classification". *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 1, no. 4. December 2017.** *The supplementary material for this paper explain how we processed the labels to introduce the "missing label information".	Accepted version Supplementary