



Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset for room level indoor localization using a smart home in a box



Ryan McConville*, Dallan Byrne, Ian Craddock,
Robert Piechocki, James Pope, Raul Santos-Rodriguez

University of Bristol, United Kingdom

ARTICLE INFO

Article history:

Received 20 March 2018

Received in revised form

8 January 2019

Accepted 18 January 2019

Available online 19 January 2019

Keywords:

Localization

RSSI

BLE

Machine learning

ABSTRACT

An annotated dataset of measurements obtained using the EurValve Smart Home In a Box (SHiB) rehabilitation monitoring system is presented. The SHiB is a low cost and easily deployable kit designed to collect data from a wrist-worn wearable in a home environment. The data presented is intended to evaluate room level indoor localization methods. The wearable device registers tri-axial accelerometer measurements which are sampled and transmitted as the payload of a Bluetooth Low Energy (BLE) packet. Four receiving gateways, each placed in a different room throughout a typical residential house, extract the accelerometer data and determine a Received Signal Strength Indicator (RSSI) for each received BLE packet. RSSI values can represent propagation losses due to distance or shadowing between the wearable transmitter and the gateway receiver.

The dataset is presented in two parts. The first is composed of four calibration or training sequences, carried out by ten participants to offer ground truth calibrations for four rooms in the house. We refer to the calibration phase as the steps taken to gather training data. The calibration procedure was designed to be as straight-forward as possible, to allow a participant to adequately train the SHiB system without supervision. Ten participants each carried out a straight forward calibration procedure once, with four participants carrying out the calibration twice, on different occasions. One participant carried out the calibration on a third occasion.

The second part of the data consists of a free-living experiment that was carried out over a period of five and a half hours starting at

* Corresponding author.

E-mail addresses: ryan.mcconville@bristol.ac.uk (R. McConville), dallan.byrne@bristol.ac.uk (D. Byrne), ian.craddock@bristol.ac.uk (I. Craddock), r.j.piechocki@bristol.ac.uk (R. Piechocki), jpope3@montevallo.edu (J. Pope), enrsr@bristol.ac.uk (R. Santos-Rodriguez).

<https://doi.org/10.1016/j.dib.2019.01.040>

2352-3409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

7.37 a.m. Of this, one and a half hours of measurements are recorded within a room containing a gateway, where one participant carried out activities of daily living while their ground-truth location was accurately annotated within each room with a gateway present. The calibration data can be used as a training scheme and the living data as a test scenario.

The dataset can be found at <https://github.com/rymc/a-dataset-for-indoor-localization-using-a-smart-home-in-a-box>

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications table

Subject area	Computer Science
More specific subject area	Internet of Things (IoT)
Type of data	Each calibration (and the test file) is an individual CSV file consisting of the fields found in Table 1.
How data was acquired	EurValve Smart Home in a Box deployed in a typical residential house.
Data format	Raw data format as recorded by the EurValve system. Data columns are then filtered to remove unrequired field such as wearable battery level, and add the ground truth labels. Further, data records where the participant was not in one of four rooms with a gateway were removed.
Experimental factors	While all calibrations were carried out within the same residential house, not all calibrations were carried out within the same period of time. For example, some calibrations may have been carried out within the same afternoon, while others on a different day.
Experimental features	10 different individuals carried out the calibration procedure within a residential house. 4 of the individuals carried it out 2 times in total. 1 of the individuals carried it out 3 times in total. 1 of the individuals then carried out free living activities within the house.
Data source location	A typical residential house in Bristol, UK (SPHERE house).
Data accessibility	Data is with this article and available online at https://github.com/rymc/a-dataset-for-indoor-localization-using-a-smart-home-in-a-box
Related research article	R. McConville, D. Byrne, I. J. Craddock, R. Piechocki, J. Pope, and R. Santos-Rodriguez, 'Understanding the Quality of Calibrations for Indoor Localisation' in <i>IEEE 4th World Forum on Internet of Things (WF-IoT 2018)</i> , Feb. 2018.

Value of the data

- This dataset consists of a number of quick and realistic training calibration scenarios. Often many indoor localization schemes using low-power channel readings, like RSSI, require labor intensive training schemes to verify their models [1–3]. This dataset contains a fast and simple calibration scheme designed deployment with a Smart Home in a Box (SHiB) system. This dataset contains RSSI data captured by ten participants following a SHiB calibration procedure. It also contains data from four participants who carried out the calibration a second time, and data from a participant who carried out the calibration process on three separate occasions. Each calibration contains the RSSI values received at each gateway for each packet received. This provides to the scientific

community a set of calibrations that meet the criteria of being quick and simple to carry out, an essential step towards wider adoption of indoor localization technology.

- The data was recorded in a residential home and is augmented with annotated location labels to denote the user room location. Video recordings with timestamps were used to match the accelerometer readings and RSSI to the participant location. The video annotation system allowed for the recording of extended living data. This provides a very reliable source of true room locations.
 - We provide a test evaluation dataset of a free-living nature, which consists of 74818 samples of a participant carrying out typical activities of daily living in a house. This data was recorded by a participant while they were staying within a typical residential house, specifically the SPHERE house [4]. They performed several daily activities that a home occupant would regularly carry out. Aside from locating an individual, these activities may be important for characterizing their behavior within the home. The addition of the activities provide an opportunity to explore the relationship between wearable BLE RSSI localization and activity.
 - This data can serve as a benchmark and motivation to encourage further research into simple calibration processes, and methods for indoor localization in smart homes. While the authors have exploited the RSSI in a previous study [5], there is potential for improving the indoor localization performance by complementing their methods with the given accelerometer data, or other methods of evaluating the reliability of, or combining, the calibrations, such as via data programming.
-

1. Data

The data is provided in a tabular format, where at every epoch the timestamp, the RSSI value, the sequence number of the packet sent by the wearable, 5 accelerometer samples, receiving gateway node, the location label corresponding to the true room and associated activity (if applicable) are described. The accelerometer readings, denoting the x , y and z axis measurements are sampled at 20 Hz from the wrist worn device which contains a tri-axial ADXL362 accelerometer with ± 4 g range [6].

The RSSI value is a relative measurement of the power in a received radio signal. In this data, the radio signal is broadcast via Bluetooth Low Energy (BLE) from the wrist worn wearable [6,7] and is potentially received by the gateways present in the house. The participant ID is denoted by the filename, 1 through 10. The instance of the calibration is also included, for e.g. if it was first, second, or third time a participant had carried out the calibrations. Thus, the final filenames are of the format 1-1.csv if it was the first instance of participant 1 carrying out the calibration, or 1-2.csv if it was the second instance of participant 1 carrying out the calibration, and so on.

A free-living dataset is also provided as a test set which was collected by a participant over a time window of approximately 5.5 h in the SPHERE house [4]. We note that we only included timestamps where the location was annotated from the video and the participant was in a room with a gateway. The final labelled dataset consists of around 1.5 h of relevant measurements. An example record can be seen in Table 1. A notable difference between the training and test data is that no activity information is recorded, as only the current position of the participant in the room was automatically recorded. However, as with the training calibration data the x , y and z accelerometer values are provided; thus the opportunity for the activity to be predicted exists.

2. Experimental design, materials and methods

This dataset is focused on the problem of localization in smart homes and thus was deployed in a somewhat typical home; a two-bedroom, two-story terraced house in a residential area. The data was recorded using the EurValve SHiB system [8]. An image of the EurValve kit can be seen in Fig. 1. The kit consists of the following:

Table 1

An explained example of a data record from the dataset.

Column	Meaning	Example
timestamp	The timestamp at which the packet was received at the gateway.	2017-07-21 07:37:46.523300,
rssi	The RSSI of the packet.	–98
seqno	The sequence number of the packet from the wearable.	4168334
sNx, sNy, sNz	The N ([1.5]) x, y and z samples from the accelerometer.	0,0.34375,0.40625
gateway	The gateway at which the packet was received.	bedroom
true_room	The true room the participant was in.	living
activity (only in training)	The activity which the participant was carrying out.	sitting

**Fig. 1.** The EurValve Smart Home in a Box (SHiB) kit used for the data collection.

- One 4G router for transmitting data from the home to a central server for analysis.
- Four gateways, each deployed in a different significant room in the home.
- One wearable that is placed on the wrist of the participant.

The SHiB kit was deployed as the intended users are instructed to by the manual; a gateway is placed in the living room, kitchen, bedroom and the upper staircase landing. The floor plans of the SPHERE house can be seen in Fig. 2.

RSSI is measured at each EurValve gateway upon packet reception. RSSI is obtained via the BLUEZ driver and using the Broadcom BCM43438 integrated Wi-Fi and BLE 4.1 SoC. Figs. 2 and 3 show the distribution of the RSSI values in each room during the calibration for two different calibrations, with a Kernel Density Estimate (KDE) fitted. This shows the volatility of the RSSI values over time. Correspondingly, Fig. 4 plots the distribution of RSSI values, with a fitted KDE, of the localized ground truth from the test data.

For dataset one, the calibration process can be described as:

- The participant places their wrist wearable very close to the gateway in order to generate a spike in the recorded RSSI. This is to facilitate the automated extraction of the RSSI values.
- Following this the participant is instructed to carry out the room specific activity for two minutes.

These steps are performed differently at each location. First the participant places the wearable as close to the gateway as possible. We include this step in the calibration process to facilitate calibration

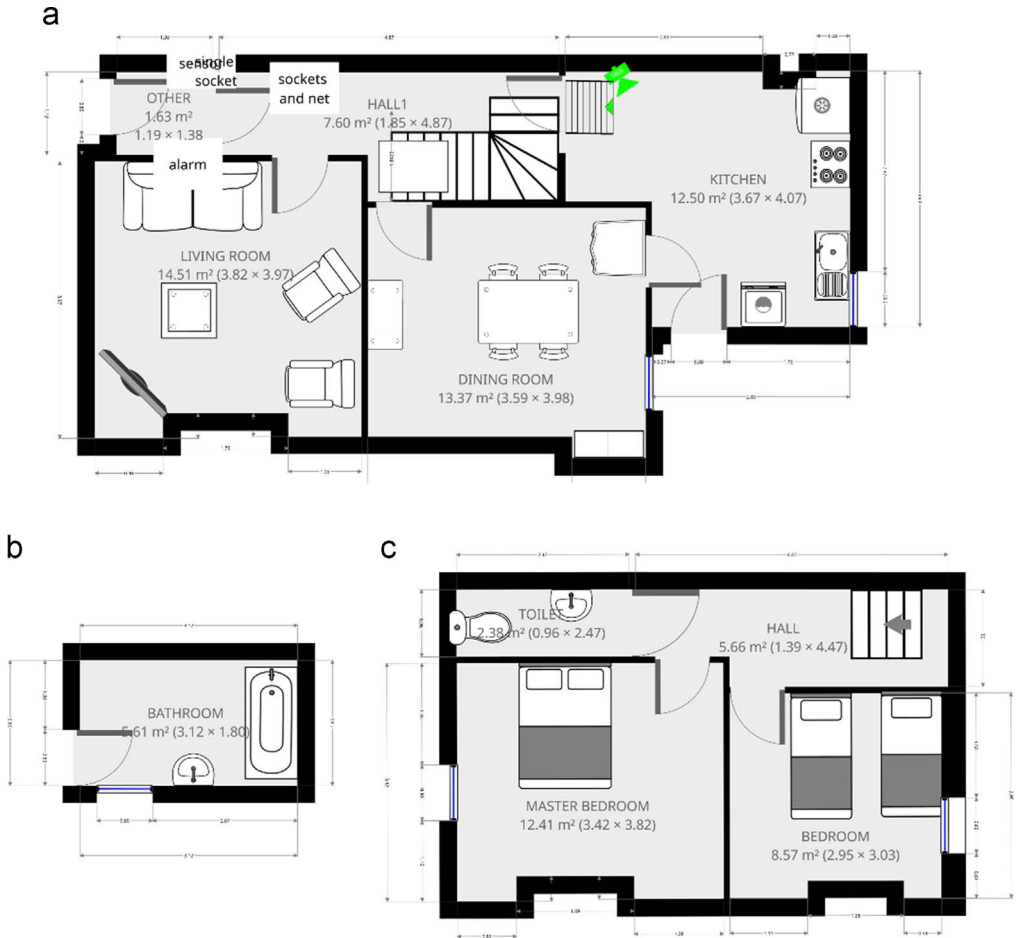


Fig. 2. Floor plan of the SPHERE house with (a) ground floor, (b) first floor bathroom and (c) second floor.

undertaken at any point after installation. By placing the wearable close to each gateway for ten seconds, it causes a large spike in the RSSI, which creates four landmarks at known intervals. Given a large amount of data, these landmarks create a known pattern for automated extraction of calibrations from large amounts of free living data. After placing the wearable beside the gateway, the participant: sits on a chair in the living room, walks around the kitchen, lays on the bed in the bedroom, and stands (sometimes, not completely stationary) by the stairs near the first floor bathroom in Fig. 5(b).

We note that the calibration process collects labels for both indoor room level localization as well as activities. Indoor localization is home dependent, that is, data collected in one home is not useful in indoor localization in different home. Thus, location labelling is the primary focus of the calibration process as it is important for high indoor localization performance. However, activity recognition does not share this property as activity labels collected for one person can be used to predict activities on another person. Nonetheless, as each person carrying out an activity does so in their own identifiable way [9], we include a step to collect activity labels in our calibration process with the aim of improving activity recognition performance on an individual basis by using this information. Fusion of localization and activity recognition has been shown to be beneficial for indoor localization [10].

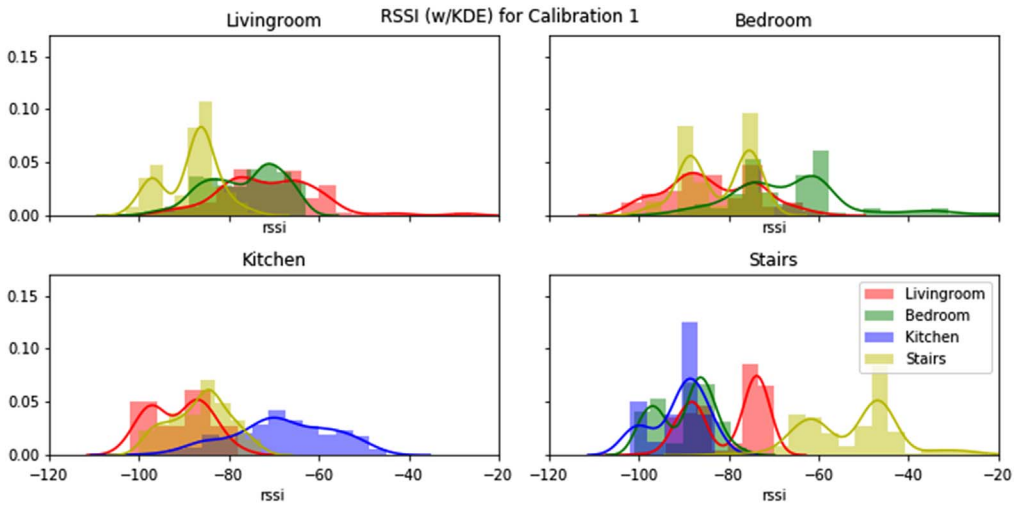


Fig. 3. RSSI values for one participant's calibration.

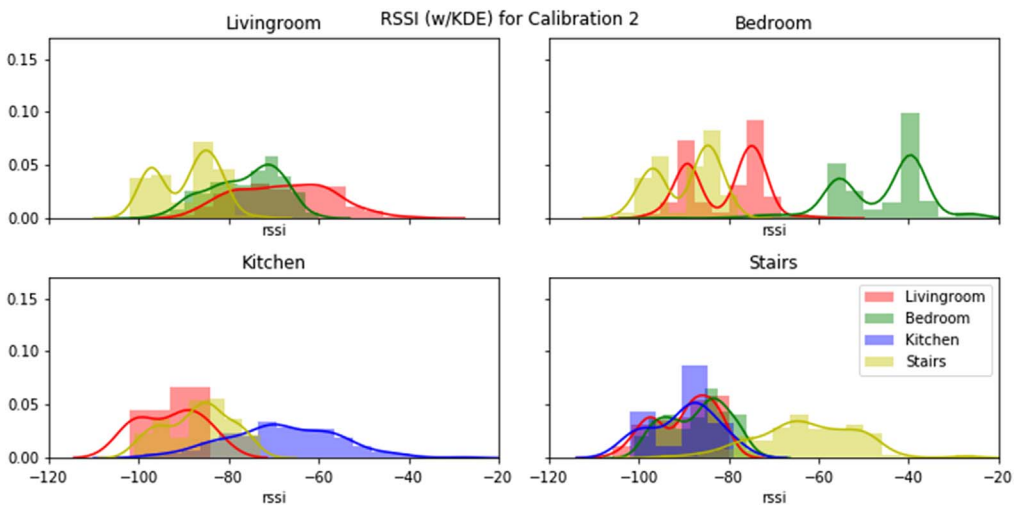


Fig. 4. RSSI values for a second participant's calibration.

The calibration process was explained to each participant who were then observed to ensure consistency between the various calibrators. Some participants who took part in the data collection had previous knowledge and experience of the calibration process, but the majority did not. Location and activity labels for the calibration data were automatically segmented and labelled based on the calibration process. It is possible for some noise in the labels if the participant did not follow the calibration procedure completely accurately; for e.g. if they carried out the activity in the room for less than the two minutes or the stood still occasionally when walking.

The free living experiments took within a single day by a single participant. This participant also carried out two calibrations (see files 2-1.csv and 2-2.csv). The participant was encouraged to mimic a daily routine and carry out tasks such as lying in bed, washing teeth in the bathroom, preparing breakfast and coffee in the kitchen, sitting on the sofa watching TV and eating a meal or working with a laptop at the dining table. A video was recorded using a camera attached to the participant. Prior to

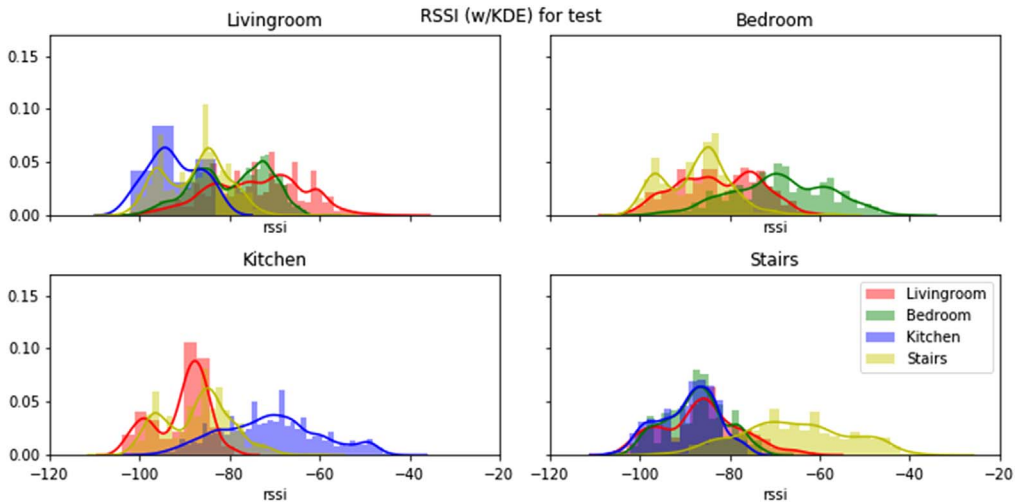


Fig. 5. RSSI values for the free living test data.

starting an experiment, the user accesses a webpage that displays the EurValve SHiB UTC time. Once recording is completed, the user must again capture the system timestamp. Both timestamps are required to mitigate any drift between the camera clock and the SHiB system. This video was later annotated with room locations and timestamps were synced to the central SHiB gateway UTC time.

Acknowledgements EurValve and SPHERE

This work was supported in part by EurValve (European Union funding) (Personalised Decision Support for Heart Valve Disease), Project Number: H2020 PHC-30–2015, 689617. This work was also supported by the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.040>.

References

- [1] D. Lymberopoulos, J. Liu, The microsoft indoor localization competition: experiences and lessons learned, *IEEE Signal Process. Mag.* 34 (5) (2017) 125–140.
- [2] S. He, S.H.G. Chan, Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons, *IEEE Commun. Surv. Tutorials* 18 (1) (2016) 466–490 (First quarter).
- [3] A. Yassin, et al., Recent advances in indoor localization: a survey on theoretical approaches and applications, *IEEE Commun. Surv. Tutorials* 19 (2) (2017) 1327–1346 (Second quarter).
- [4] N. Zhu, et al., Bridging e-health and the internet of things: the SPHERE project, *IEEE Intell. Syst.* 30 (4) (2015) 39–46.
- [5] R. McConville, D. Byrne, I. Craddock, R. Piechocki, J. Pope, R. Santos-Rodriguez, Understanding the Quality of Calibrations for Indoor Localisation in IEEE 4th World Forum on Internet of Things (WF-IoT 2018), Feb. 2018.
- [6] X. Fafoutis, et al., Designing wearable sensing platforms for healthcare in a residential environment, *EAI Endorsed Trans. Pervasive Health Technol.* 17 (12) (2017).
- [7] S. Dumanli, L. Sayer, E. Mellios, X. Fafoutis, G.S. Hilton, I.J. Craddock, Off-body antenna wireless performance evaluation in a residential environment, *IEEE Trans. Antennas Propag.* 65 (11) (2017) 6076–6084.

- [8] J. Pope, et al., 'SPHERE in a Box: Practical and Scalable EurValve Activity Monitoring Smart Home Kit', in 2017 IEEE in: Proceedings of the 42nd Conference on Local Computer Networks Workshops (LCN Workshops) pp. 128–135, 2017.
- [9] R. McConville, R. Santos-Rodriguez and N. Twomey, Person identification and discovery with wrist worn accelerometer data in Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2018.
- [10] M. Kozłowski et al., Data Fusion for Robust Indoor Localisation in Digital Health in IEEE Wireless Communications and Networking Conference Workshops (WCNCW 2018).