data

MDPI

Data Descriptor

Multi-Modal Dataset of Human Activities of Daily Living with Ambient Audio, Vibration, and Environmental Data

Thomas Pfitzinger [1,*], Marcel Koch [1,2], Fabian Schlenke [1] and Hendrik Wöhrle [1]

[1] Institute of Communication Technology, Department of Information Technology, Dortmund University of Applied Sciences and Arts, Sonnenstraße 96, 44139 Dortmund, Germany; marcel.koch@materna.group (M.K.)
[2] Materna Information & Communications SE, Robert-Schuman-Straße 20, 44263 Dortmund, Germany
* Correspondence: thomas.pfitzinger@fh-dortmund.de

Abstract: The detection of human activities is an important step in automated systems to understand the context of given situations. It can be useful for applications like healthcare monitoring, smart homes, and energy management systems for buildings. To achieve this, a sufficient data basis is required. The presented dataset contains labeled recordings of 25 different activities of daily living performed individually by 14 participants. The data were captured by five multisensors in supervised sessions in which a participant repeated each activity several times. Flawed recordings were removed, and the different data types were synchronized to provide multi-modal data for each activity instance. Apart from this, the data are presented in raw form, and no further filtering was performed. The dataset comprises ambient audio and vibration, as well as infrared array data, light color and environmental measurements. Overall, 8615 activity instances are included, each captured by the five multisensor devices. These multi-modal and multi-channel data allow various machine learning approaches to the recognition of human activities, for example, federated learning and sensor fusion.

Dataset: DOI: 10.5281/zenodo.7937591, direct download link: https://zenodo.org/api/records/7937591/files-archive (accessed on 13 August 2024).

Dataset License: CC-BY (v. 4.0)

Keywords: sensor data; labeled data; internet of things; smart home; assisted living; human activity recognition; machine learning; classification; activities of daily living

Citation: Pfitzinger, T.; Koch, M.; Schlenke, F.; Wöhrle, H. Multi-Modal Dataset of Human Activities of Daily Living with Ambient Audio, Vibration, and Environmental Data. *Data* **2024**, *9*, 144. https://doi.org/10.3390/data9120144

Received: 13 August 2024
Revised: 28 November 2024
Accepted: 3 December 2024
Published: 9 December 2024

1. Summary

Detecting activities of daily living (ADLs) is relevant for healthcare monitoring, smart home automation, security, and energy management [1,2]. Knowledge of these activities allows us to better understand the context to base automated decisions on. A list of published human activity datasets is shown in Table 1.

Many datasets are based on wearable inertial sensors [3–9]. Increasingly more people are carrying a smartphone or smartwatch on them, which usually has such sensors integrated. Using this type of data has the benefit that activities can be tracked, regardless of the subject's location. However, when the carried device is removed, activity recognition is no longer possible. For this reason, ambient sensors can be a preferable alternative, especially in scenarios where subjects cannot or do not want to carry wearable devices continuously.

Ambient sensors are fixed in the environment to capture the user and their interaction. Inertial sensors have been applied in this way [9]. In a different approach, measurements of stationary radars [10] and Wi-Fi signals have been recorded [11,12]. These methods can identify the position and movement of subjects. Other datasets capture the sound of the activities with a single stationary microphone [13,14]. Mudharanga et al. utilized three

microphones placed at various distances and included depth video recordings, which can capture poses and movement [15].

Table 1. Comparison of related human activity datasets.

Dataset	Year	Sensor Placement	Type of Data	No. of Activities	No. of Participants
Shoaib et al. [3]	2014	Wearable	Various	7	10
WISDM [4,16]	2011, 2019	Wearable	Acceleration	18	51
Garcia-Gonzalez et al. [5]	2020	Wearable	Acceleration	4	19
Climent-Pérez et al. [6]	2022	Wearable	Acceleration	24	52
Matey-Sanz et al. [7]	2023	Wearable	Acceleration	5	23
UESTC-MMEA-CL [8]	2024	Wearable	Video and acceleration	32	10
Opportunity [9]	2010	Ambient and wearable	Acceleration	17	4
Narayanan, Zenaldin [10]	2015	Ambient	Radar	18	6
Alsaify et al. [11]	2020	Ambient	Wi-Fi	12	30
Alazrai et al. [12]	2020	Ambient	Wi-Fi	12	66
Stork et al. [13]	2012	Ambient	Audio	22	–
Siantikos et al. [14]	2017	Ambient	Audio	5	–
Madhuranga et al. [15]	2021	Ambient	Audio, video	24	17
Proposed dataset	2024	Ambient	Audio, Vibration, infrared array, light color, environmental	25	14

The dataset presented in this paper focuses on ambient sensors, allowing us to classify ADLs non-intrusively. While this requires a certain proximity of the subject to the installed sensor, it is not dependent on the subject carrying a sensor device on their body. The dataset can facilitate human activity recognition, which provides a basis for different applications ranging from assisted living to home automation, security and energy management. It is distinctive in that five identical sensor devices were used, and multiple modalities were recorded: audio, vibration, infrared array, light color, and environmental data.

The proposed dataset offers a large body of 8615 human activity recordings totaling about 17.5 h. The activities were performed by a one participant at a time. Each recording was captured by five identical sensors at different locations. This makes it possible to use the data as a five-channel time-series for a central model or for federated learning approaches. Alternatively, the data can be merged into a single channel, increasing the number of items by a factor of five. With 25 different activity classes, a wide variety of common movements and interactions in a household are covered. Examples are walking, sweeping the floor, typing, or opening or closing a door. These short atomic activities can serve as a basis for the recognition of more complex activities, pattern recognition or scene classification.

The dataset has been applied for the recognition of human activities on an embedded system using audio [17]. A subset of the activity classes is used to train a convolutional neural network with audio spectrograms as input. Different options for reducing the model complexity are explored in order to find a model variant that can run on the resource-constrained system.

2. Data Description

The dataset consists of 25 different atomic ADLs, including basic human movements (walking, sitting down, standing up) and interactions with various objects and appliances

(door, refrigerator, window, light, etc.). They were recorded in an environment with a kitchen and meeting room at the Dortmund University of Applied Sciences and Arts. The data consist of audio, three-axis vibration, an infrared array, light color and environmental measurements. All activities were captured by five identical multisensors distributed across the two rooms. This results in five dataset entries for each activity instance that can be used in combination or independently of each other.

2.1. Activity Classes

The list of ADLs performed with their corresponding total duration is shown in Table 2. With *No activity* containing the most data, the imbalance ratio to the least represented class (*Light off*) is 10:133. The over-represented classes can be undersampled to reduce this imbalance. After undersampling *No activity* to match the next largest class (*Walk*), the imbalance ratio is reduced to 10:42. This remaining imbalance can be countered by using class weights when training a model. If deemed necessary, oversampling or the augmentation of the less represented classes can be considered.

Table 2. Recorded activities, their total duration, and their description. The total duration is the sum of all activity instances, which each consist of five recordings from the five multisensors.

Activity	ID	Number of Items	Accumulated Duration [h]	Portion of Total Duration [%]	Description
No activity	0	1155	19.23	22.01	No person is present in the recording environment
Walk	1	2320	6.05	6.93	Walking from any one point to another, possible passing the open door
Walk to room	2	1865	5.49	6.29	Walking from any point in one room through the open door into the other room
Open door	3	2240	2.61	2.99	Opening the door between living room and kitchen
Close door	4	2270	2.89	3.31	Closing the door between living room and kitchen
Open window	5	2325	2.79	3.20	Opening a window in the living room
Close window	6	2325	2.68	3.07	Closing a window in the living room
Sit down	7	2330	3.69	4.22	Pull up a chair and sit down
Stand up	8	2300	3.90	4.47	Get up from a chair and push or place the chair under the table
Light on	9	1920	1.95	2.23	Press a push-button, turning on the light in the living room
Light off	10	1845	1.45	1.66	Press a push-button, turning off the light in the living room
Typing	11	70	3.88	4.44	Typing random sentences on a keyboard placed on the table in the living room
Make coffee	12	360	3.41	3.90	Make a coffee using the Senseo coffee machine in the kitchen; does not include the heating process of the water
Place plate	13	1850	1.53	1.75	Place a plate on the table in the living room

Table 2. *Cont.*

Activity	ID	Number of Items	Accumulated Duration [h]	Portion of Total Duration [%]	Description
Sweep	14	45	2.55	2.92	Sweep the floor of both rooms with a broom
Vacuum cleaning	15	70	3.85	4.40	Vacuum the floor of both rooms
Start toaster	16	1865	1.55	1.78	Push down the switch to start the (empty) toaster in the kitchen
Stop toaster	17	1840	1.47	1.69	Press the button to release and stop the (empty) toaster in the kitchen
Open fridge	18	1775	1.79	2.05	Open the refrigerator in the kitchen
Close fridge	19	1795	1.99	2.28	Close the refrigerator in the kitchen
Open dishwasher	20	2265	2.25	2.58	Open the dishwasher in the kitchen
Close dishwasher	21	2265	2.34	2.68	Open the dishwasher in the kitchen
Place on stove	22	1900	1.63	1.87	Place a pot on the stove in the kitchen
Take from stove	23	1840	1.57	1.80	Remove a pot from the stove in the kitchen
Wipe stove	24	2240	4.82	5.51	Wipe the stove in the kitchen with a dry or wet cloth

The activities themselves differ in duration (see Figure 1). Most activities are short with a single action being performed, for example, closing a door. The median length of these short recordings is between 2.7 and 10.7 s. The activities *Vacuum cleaning*, *Sweeping*, and *Typing* are continuous in nature. These recordings are longer, with a duration of about 200 s. The length of *Make coffee* is determined by the coffee machine and has a median duration of 32 s. *No activity* recordings were taken in 60 s time windows.

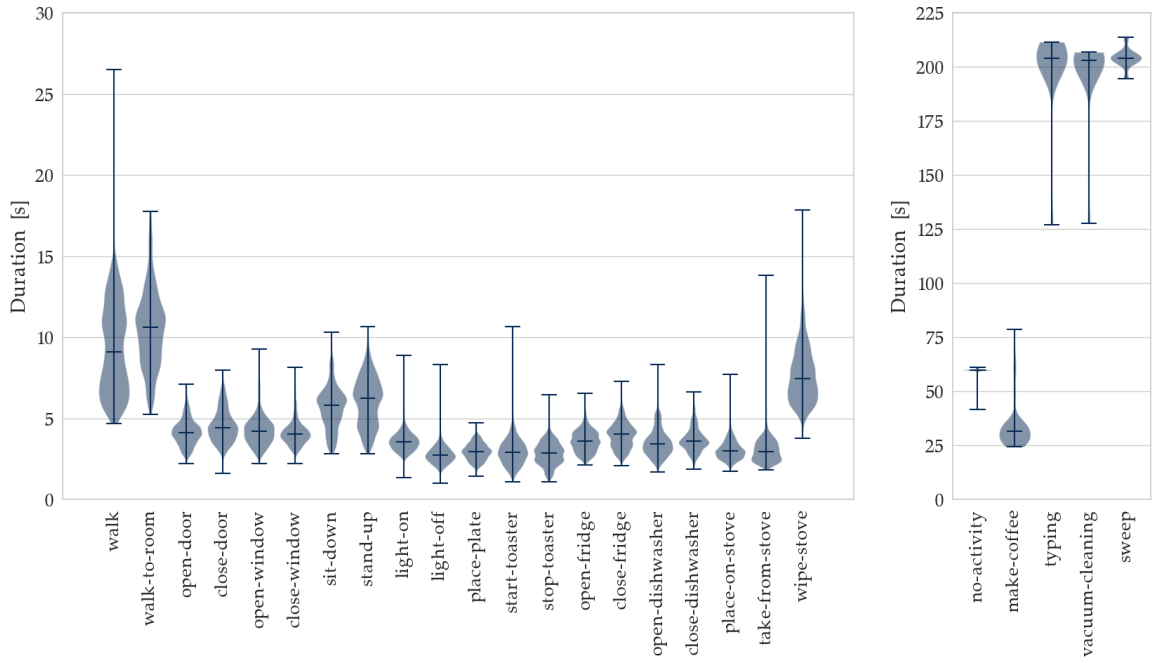

Figure 1. Distribution of activity instance duration within each class, as well as the minimum, median, and maximum duration. Separated into short and long activities.

2.2. Types of Data

The main volume of the dataset consists of ambient audio and three-axis vibration recordings. With a high data rate, these data potentially provide enough information to recognize all activities. Infrared array, light color, and environmental data recorded with low sample rates can contain additional useful information.

As can be seen in Table 3, audio was sampled at 16 kHz and a vibration at 1.085 kHz. While the audio sample rate is constant, the vibration sample rate deviates by up to 10 Hz. For this reason, the dataset includes the vibration sample rate for each entry. The infrared array and light color data were recorded with 1 Hz. The environmental was recorded once every three seconds. As these measurements are mostly constant over short time spans, only an average value is included for each activity, instead of multiple values. Therefore, the recording sample rate of 0.33 Hz does not correspond to the sample rate in the dataset.

Table 3. Recorded data types, their sample rate and format.

Type of Data	Sample Rate	Format of Single Sample	Range	Unit/Description
Audio	16 kHz	Integer value	−32,768 to 32,767	Raw; mono-audio amplitude
Vibration	1.085 kHz	Array of 3 integer values	−32,768 to 32,767 corresponding to −2.5 to 2.5 g [1]	*x*-, *y*- and *z*-axis
Infrared array	1 Hz	8 × 8 matrix of float values	−65 to 300	Matrix of temperature in °C
Light color	1 Hz	Array of 4 integer values	0 to 65,536	Raw; red, green, blue and clear
Temperature	0.33 Hz (single measurement per entry)	Floating-point value	−65 to 300	°C
Relative humidity		Floating-point value	0 to 100	%
Atmospheric pressure		Integer value	0 to 20,000	daPa [2]
Air quality index		Integer value	0 to 500	See [18]
VOC		Integer value	0 to 10,000	ppm
CO_2 equivalent		Integer value	0 to 10,000	ppm

[1] g = 9.81 ms^{-2}; [2] daPa = 10 Pa = 10^{-1} hPa.

Three examples of activities recorded by one multisensor are shown in Figure 2. The high-frequency audio and vibration data are continuous over the duration of the element, while light and infrared-array values are represented by singular points. The environmental data are not included in the figures, as the values do not change over one activity instance. Each infrared array measurement is an 8 × 8 matrix represented by the mean value. The complete infrared array data for the displayed recording of *Walk to room* is shown in Figure 3. When the participant walks in front of the sensor, local changes in the infrared temperature readings are noticeable.

A detailed overview of the recorded data of the different sensor types can be found in Appendix A. The data distribution over each activity class is displayed, showing to what extent the different activities were captured best by each sensor.

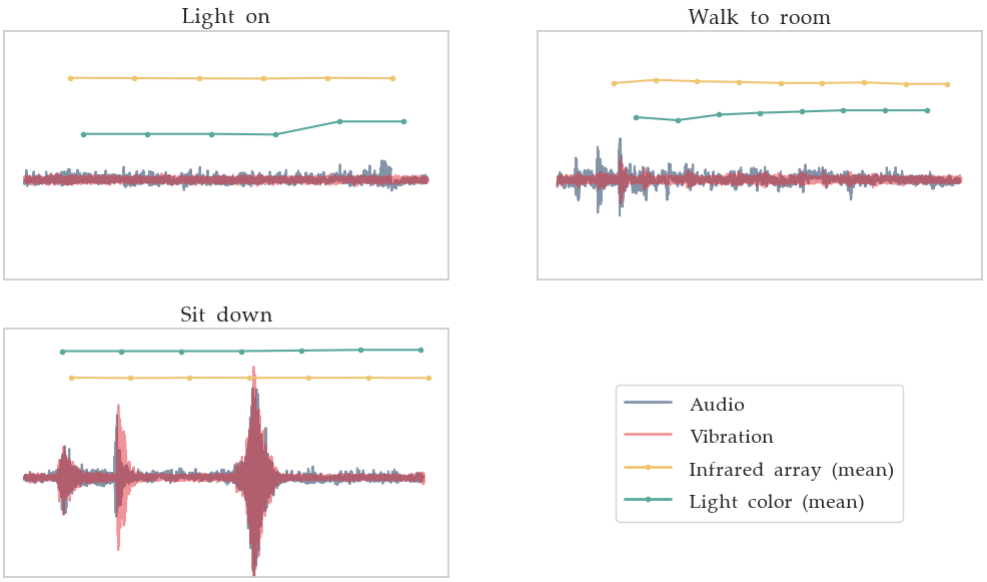

Figure 2. Three examples of activity recordings from multisensor 4. The measurements for the different environmental readings are not depicted as they each consist of a single value for the example.

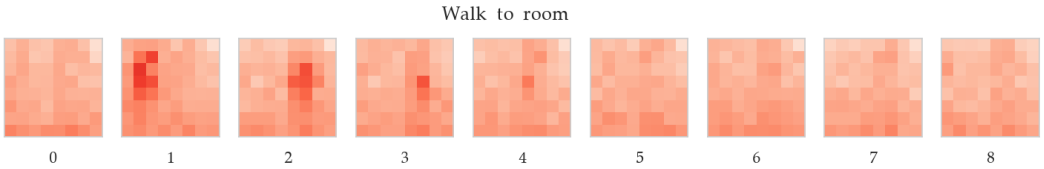

Figure 3. The infrared array data from the *Walk to room* example shown in Figure 2. The participant enters the sensor's field of view from the left and then walks away from the sensor. Each second, an 8×8 matrix of IR-temperature readings is captured, displayed as a heatmap.

2.3. Dataset Structure

The data are organized in a folder structure with one folder for each activity, as shown in Figure 4. The folder names consist of the activity ID (0–24) and name. Each folder contains one file per sensor device, in which the activity instances recorded by that device are stored. The file names start with the folder they are in, followed by the multisensor ID (1–5).

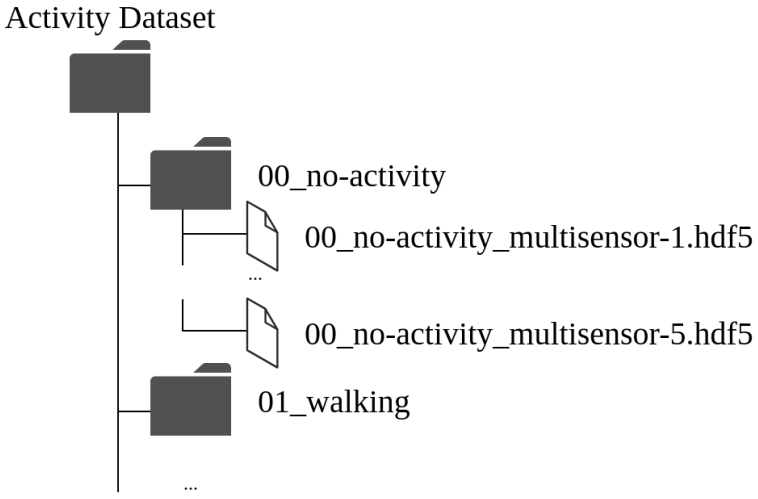

Figure 4. Folders and files in the dataset.

The file format is HDF5, allowing us to store heterogeneous data in a table structure, which is shown in Figure 5. Each data type is stored in a column, while each row represents one activity recording. The first columns contain additional metadata:

- The relative start time of the activity in seconds.
- A unique trial ID for each activity instance that is the same for the recordings across all five multisensors. It can be used to join the recordings from different multisensors for each activity instance.
- The subject ID of the activity, showing which participant performed this activity (see Figure 6).
- The sample rate for the vibration data.

The remaining data columns can be grouped into three categories:

- High-frequency data that start at the timestamp given in the *start* column and have a known data rate. Each entry in these columns can have a different number of samples depending on the length of the recording.
- Low-frequency data that do not change significantly within one activity instance and were therefore averaged to a single value per entry.
- Low-frequency data that include a timestamp for each value, which is stored in a corresponding column. These columns have multiple values per activity and therefore contain more entries than the other columns.

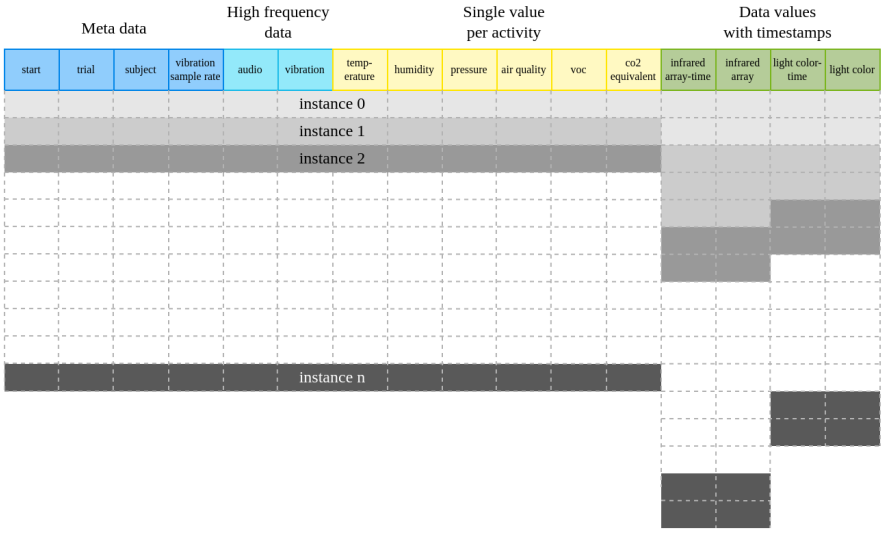

Figure 5. Table structure of the data in each HDF5-file. For one recording, the high frequency data consist of an array, and a single value is given for the environmental data. *infrared-array* and *light-color* have multiple values, each with a corresponding timestamp in the neighboring column.

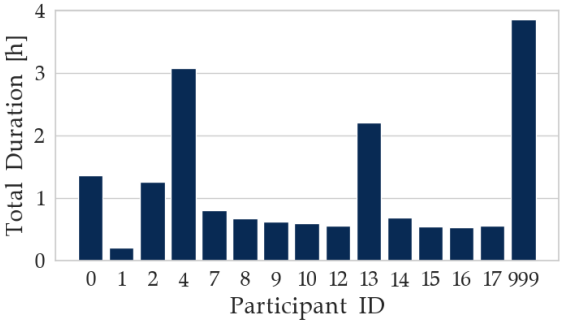

Figure 6. Participant IDs and total time of the recordings. ID 999 is used for *No activity*, where no participant was involved.

3. Methods

The dataset was recorded and labeled in supervised sessions in which the participant performed activities in the recording environment. Five multisensor devices were placed in the recording environment. They transmitted the data wirelessly to a separate server, where they were saved. Finally, the data were filtered to remove faulty records, and the different types of data were synchronized.

3.1. Participants

The activities were performed by 14 healthy volunteers (13 males and 1 female). Due to different availability, some participants contributed more to the recording than others. The participant IDs with the duration of the contributed recording are shown in Figure 6. The special ID 999 was used for *No activity*, as no participant was involved in these recordings.

All participants were informed beforehand about the utilization of the recorded data and the intent to publish them. The dataset is pseudonymized by using numerical IDs to represent the participants. Furthermore, no identifying information of the participants is contained in the dataset.

3.2. Recording and Labeling Process

The data were labeled during the recording process. The recording included the participant performing the activities and an observer who initiated and controlled the process. The observer could monitor the rooms via camera live streams. The recording of a complete set of all activities with one participant was carried out in two sessions that each lasted up to two hours.

The observer controlled the recording session using an application with a simple web interface developed for this purpose. The sequence of recording a set of activity instances is depicted in Figure 7. The observer would select an activity and click on a button to initiate the recording. With this, the server would start to save the vibration, infrared-array data, and audio streams to files. The selected activity was shown to the participant on a display in the recording environment. When the participant was ready, the observer would click on a button to start an instance of the activity. Text on a display, as well as an acoustic countdown (beeps) signaled to the participant to start executing the activity. The participant performed the activity, and, on seeing on the live video feed that it was completed, the observer would click another button to mark the stop time. The start and stop timestamps were automatically saved to a time series database along with the label of the current activity. This process was repeated 20 or 40 times in a set for one activity before stopping the recording and moving on to the next activity.

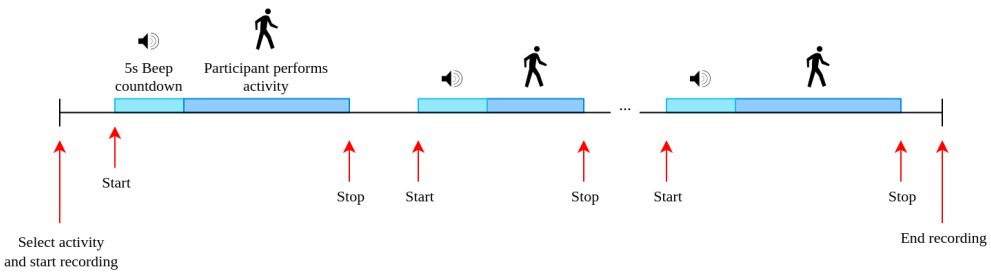

Figure 7. Recording sequence of one activity set. The red arrows represent inputs by the observer.

Each activity instance was automatically marked with a unique trial ID. If disturbances or mistakes occurred, the trial ID was logged by the observer in order to remove the corresponding recordings later.

Complementary activities, e.g., opening and closing a door or sitting down and standing up, were performed alternately within one set. The activity *Make coffee* was performed 3 or 6 times for one set, as each execution took a longer time. Similarly, the activities *Vacuum cleaning*, *Sweeping the floor*, and *Typing* were recorded in a single continuous trial of

130–200 s. This allowed a more natural execution of the activities. *No activity* was recorded during nighttime when nobody was present in the rooms.

3.3. Data Storage

Due to the different nature of the data types, there were three methods for storing the data. The sensors that produced small amounts of data were transmitted over MQTT to a time series database. These data were stored continuously, independently of the recording sessions. Vibration and infrared array data were also published with MQTT, but because of the multidimensional structure of the samples, they were stored in files on the recording server. This was only initiated during recording sessions to conserve storage space. Similarly, the audio was only activated for the recording sessions and saved in the recording server's file system. However, due to the high sample rate of 16 kHz, audio was not transmitted with MQTT, but with a direct RTP stream. Audio was stored as raw values in WAV-files, and vibration and infrared array data were stored in HDF5-files.

Vibration and infrared array data was transmitted in packets with timestamps generated by the multisensor. The audio stream, however, did not contain timestamps. Therefore, timestamps were generated by the recording server on receiving the data. In addition, the recording server performed sample rate adjustment on the incoming stream to achieve a stable sample rate of 16 kHz. This was implemented because the audio streams showed variation in their sample rate in previous recordings.

3.4. Multisensors

The multisensors used to capture the data of this dataset were developed by Insta GmbH[1] (see Figure 8a). They are plugged into an outlet (see Figure 8b) and connect to the local network wirelessly. An ESP32-microcontroller collects the data from connected sensors and transfers them using MQTT, or RTP in the case of audio data. The specific sensors used to record the data are shown in Figure 8c and listed in Table 4.

(**a**) (**b**) (**c**)

Figure 8. Front and side views and the internal circuit board of the multisensors used for recording the data. (**a**) Multisensor front view. (**b**) Multisensor side view. (**c**) Circuit board of a multisensor. ESP32 (A), microphone (B), accelerometer (C), infrared array (D), light color sensor (E), environmental sensor (F).

The multisensors synchronized their local clock via Network Time Protocol (NTP). Recorded data points, except for audio, were provided with a Unix epoch timestamp of the time they were captured. The data points and timestamp were packaged in JSON format and published over MQTT. As the vibration data had a higher data rate, the data points were not published individually but in bundles of 110 samples with the timestamps for the first captured sample. No timestamps were added to the audio samples, due to the high sample rate and the transfer with RTP instead of MQTT and JSON.

Table 4. Sensors used for the data recording.

Data Type	Sensor	PCB Label (Figure 8c)
Audio	IMP34D (STMicroelectronics)	B
Vibration	LIS3DHH (STMicroelectronics)	C
Infrared array	AMG88 (Panasonic)	D
Light color	TCS3472 (AMS)	E
Temperature Relative humidity Atmospheric pressure Air quality measure VOC CO_2 equivalent	BME680 (Bosch)	F

3.5. Recording Environment

The experimental environment consisted of two rooms. The larger room is a dining or meeting room with a large table, multiple chairs, two wall-mounted displays, and some cupboards. The second and smaller room is a kitchen equipped with typical appliances like a stove (ceramic hob), sink, refrigerator and dishwasher. A toaster and Senseo coffee machine were also present and were used for the recording. Figure 9 shows the room layout with the position and orientation of the five multisensors. The multisensors were plugged into wall outlets or power strips, always facing into the room.

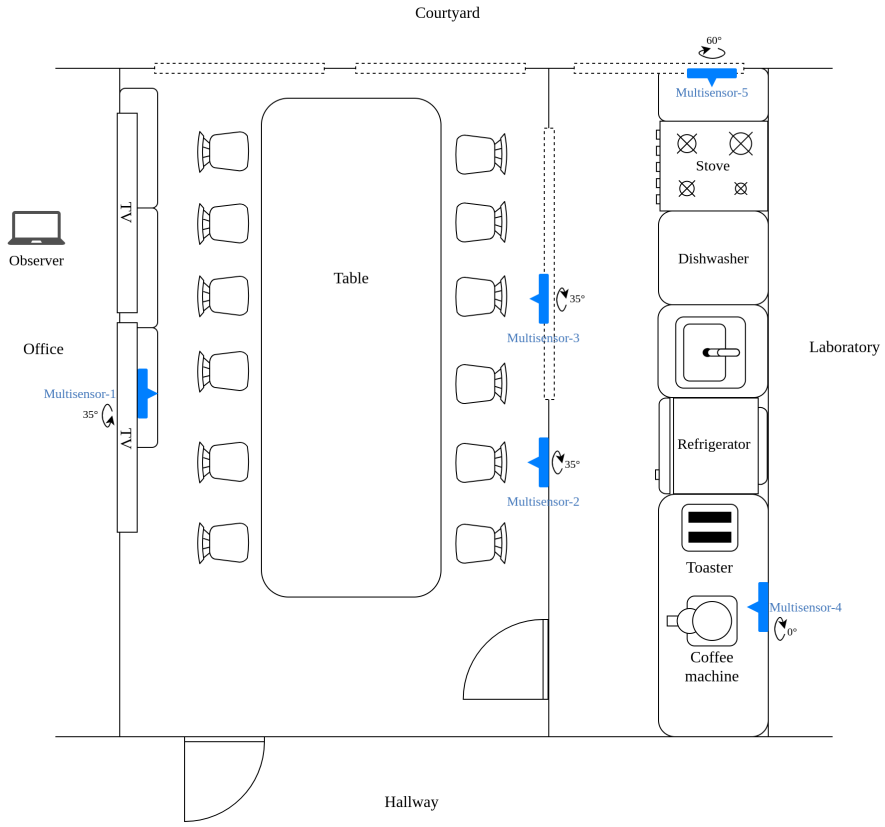

Figure 9. Layout of the two rooms that composed the recording environment. The multisensor positions are marked with blue rectangles and a triangle pointing to the direction they are facing. The rotation of the sensors along the facing axis is also included.

The windows and the entrance door were kept closed during recording, and the door connecting the two rooms was kept open. Exceptions from this are the activities involving a window and door, for which one of the courtyard-facing windows and the connecting door were used.

A camera was installed in each room to allow the observer to monitor the recording sessions without being present in the room. The video data also served for later analysis and verification of the data but are not published with the dataset for privacy reasons.

3.6. Analysis and Filtering

3.6.1. Removing Faulty Activity Instances

The recorded audio files were split into individual activity instances, as each file included a complete set of activities (see Figure 7). This was carried out using the saved start and stop timestamps and by detecting the audible beep countdowns through frequency matching as a supplementary measure.

After removing instances marked as faulty during the recording, the recorded data were analyzed in three steps to identify elements containing undesired noises or faulty data. In a first step, instances were removed where the vibration sample rate significantly deviated from the average sample rate. With this, the vibration sample rate was limited to the range between 1075 and 1095 Hz.

Secondly, the voice activity detection model Brouhaha [19,20] was applied to each audio recording. All elements where the result showed a voice probability over 75 % for at least one second were then inspected manually. Those containing audible voice or other noises were removed from the dataset.

Lastly, outliers of each activity class were inspected. To detect them, four features were calculated for each audio recording:

- Standard deviation;
- The absolute peak-to-peak range;
- Standard deviation of the frequency bins (after Fourier transform);
- The average of the frequency bins (after Fourier transform).

Based on these features, the points with the largest distance to the nearest neighbors were determined as outliers. These recordings were then manually inspected, and those containing undesired noise were removed.

When removing a recording, the complete activity instance was excluded from the dataset. This includes all sensor data of that activity instance from all multisensors. With this analysis, 510 activity instances were removed from the dataset.

3.6.2. Synchronization of Sensor Data

After filtering the recordings, the different types of data were then combined. The audio data streams did not provide timestamps for the time of recording. Therefore, the time of arrival of the audio on the recording server was saved for each audio file. Due to latency in the data transmission and slightly deviating sample rates, these timestamps did not entirely match those of the other data types. To better synchronize the audio with the other data, a cross-correlation between audio and vibration was performed. Many of the recorded activities showed similar spikes in the audio and vibration data. Where these similarities were sufficient, they were utilized to synchronize the recordings (see Figure 10). For this purpose, the audio was down-sampled to the vibration's sample rate. Both signals were passed through a low-pass filter, and the cross-correlation was calculated. The original audio was then shifted according to the maximum correlation.

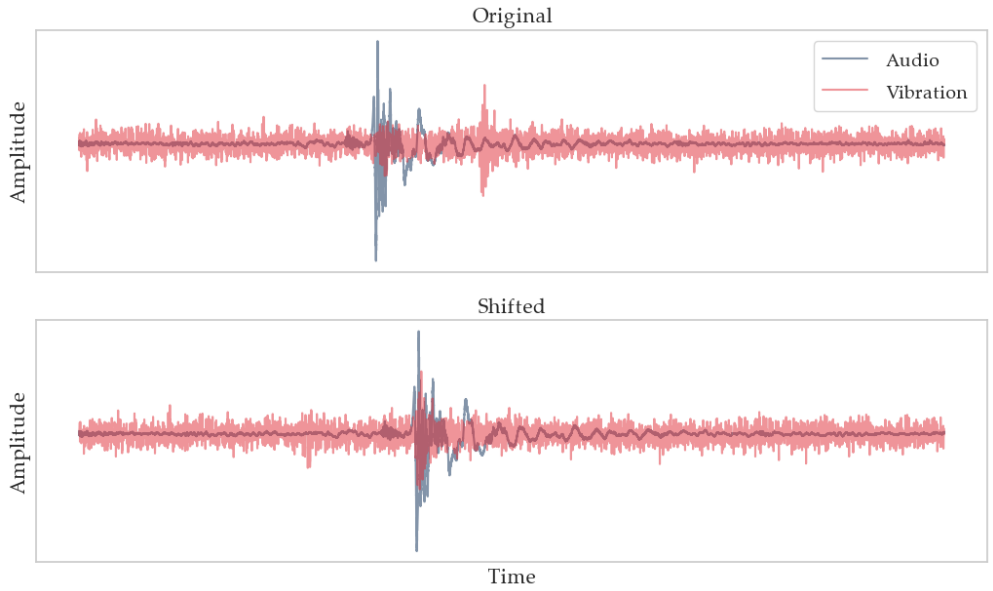

Figure 10. Audio and vibration before and after cross-correlation.

The synchronization of audio and vibration with the low-frequency data was trivial, as each of these values was saved with a timestamp. The environmental data (temperature, humidity, atmospheric pressure, air quality, VOC and CO_2) showed no significant changes over the course of a single activity instance. As a simplification, they were therefore averaged and saved as a single value for each activity instance. The infrared array and light color data were added unchanged with a timestamp included at each data point.

4. Usage Notes

4.1. Data Access

The dataset can be downloaded from the provided repository in its entirety or in parts, each part being one activity class. After unzipping the downloaded archives, the files can be read with common tools or libraries for HDF5. The dataset repository includes an example python script with which all or part of the data can be loaded. For example, a subset of activity classes and sensor types can be loaded.

4.2. Different Sensors

Note that, naturally, not all activities are captured equally well on all sensors. Most activities produce distinct sounds (e.g., footsteps, the door closing, and typing) that are captured by the microphones of all five multisensors. Exceptions to this are the activities involving the stove and light switch, which are less audible. The accelerometer is more dependent on proximity to the activity and the propagation of the vibration through the environment. Direct impacts to the floor or wall (e.g., footsteps or a door or window closing) are captured better than activities on a table (e.g., typing and placing a plate). Furthermore, some activities are only captured by the vibration sensors in the close vicinity (e.g., *Open fridge*, *Start toaster*). The perception of the infrared is confined to the area in front of the sensor. Therefore, many recordings may not contain data of the activity, as it was performed outside the sensor's field of perception. The light sensor contains relevant information of the activities *Light on* and *Light off*, but little information about other activities.

When working with the infrared array data, it is to be noted that light- and heat-emitting devices in the room and direct sunlight can cause IR artifacts. Furthermore, the device rotation plays a relevant role. It is not the same for all five devices (see Figure 9).

4.3. Limitations

Some limitations apply to the dataset and should be considered when using it:

- The participants do not represent the general population, as most were male, and detailed information about body height, weight and age cannot be provided.
- The microphone was affected by vibrations because a MEMS microphone was used. Therefore, some audio recordings have low-frequency disturbances. This is most prevalent in multisensors 2 and 3, which were mounted on the same wall, in which the door was opened and closed. The vibration in the wall caused by this is superimposed on the sound captured by the microphone. To counteract this, low frequencies can be removed from the audio. A high pass filter that cuts off frequencies below 25 Hz is sufficient for this purpose. We provide unfiltered audio to avoid loss of potentially useful information.
- While audio and vibration data are consistently available for all dataset entries, some instances are missing other measurements due to data loss. The incomplete data elements are not excluded from the dataset, because audio and vibration contain usable information. Overall, 4034 items are missing values (4019 environmental data, 2800 light color, and 15 infrared array). The affected items include all activity classes. When extending these instances to all five sensors, 1453 instances or 7265 items are affected, and removing them reduces the dataset by 15.67 %. Depending on which measurements are needed, instances that are missing the required data can be dropped. Nonetheless, the reduced dataset still provides a sufficient size to train models.
- The activity class *Walk to room* is clearly defined as walking and passing through the door. However, the class *Walk* may include walking through the door or walking only within one room. Therefore, these two classes have some overlap, and they might be difficult to distinguish. Depending on the application, it may be sensible to treat both as one class or to discard one.

Author Contributions: Conceptualization, M.K. and F.S.; methodology, M.K. and F.S.; software, T.P., M.K. and F.S.; data curation, T.P. and M.K.; writing—original draft preparation, T.P.; writing—review and editing, H.W.; supervision, H.W.; project administration, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry of Economic Affairs and Energy (BMWi)–now Federal Ministry for Economic Affairs and Climate Action (BMWK)–in the ForeSight project, grant number 01MK20004G.

Institutional Review Board Statement: Ethical review and approval were waived for this study, because no personal or identifying data of the participants was stored permanently.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in Zenodo at https://zenodo.org/records/7937591 (accessed on 4 December 2024) or https://doi.org/10.5281/zenodo.7937591 (accessed on 4 December 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ADLs	Activities of Daily Living
HDF5	Hierarchical Data Format, version 5
ID	Identifier
JSON	JavaScript Object Notation
MQTT	Message Queue Telemetry Transport
RTP	Real-time Transport Protocol
VOC	Volatile Organic Compound
WAV	Waveform Audio File Format

Appendix A

Figures A1–A10 feature the distribution of each sensor type across each activity class as a violin plot with medians. The data of all five multisensors were joined for this evaluation. Some sensors show greater variation or an offset for certain classes, indicating a higher potential to distinguish this class from others. For example, in Figure A1, the audio distribution for *Vacuum cleaning* is distinctly offset and more spread out compared to that of *Sweep*. The vibration data show high variation for *Open door* and *Close door* (see Figure A2), while the light measurements vary noticeably for the *Light on* and *Light off* activities (see Figure A4).

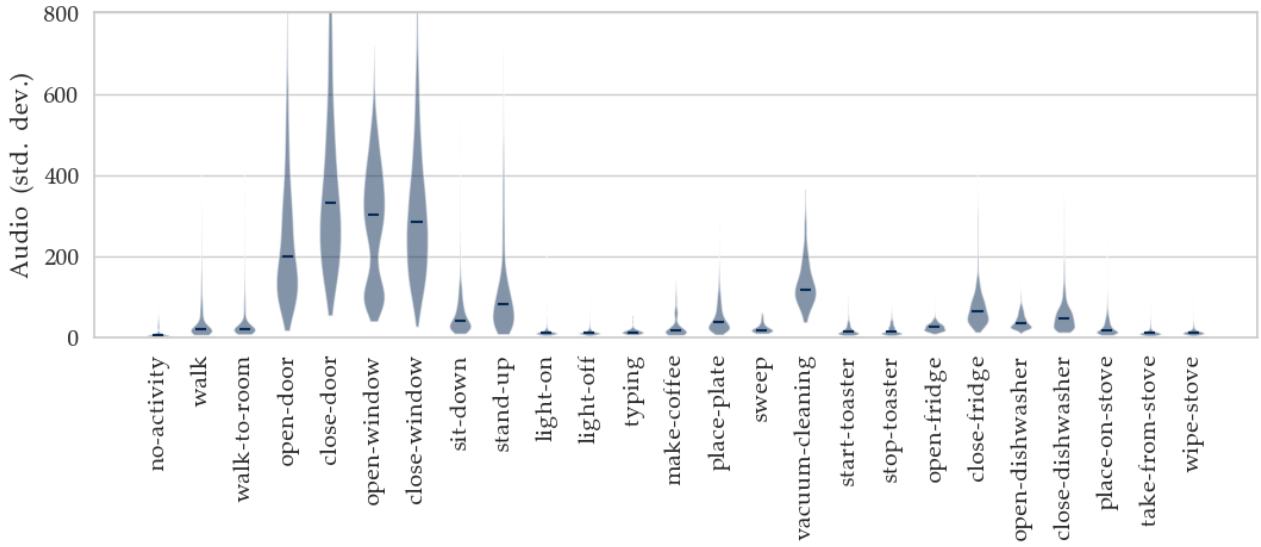

Figure A1. Variation of audio data for each activity class using the standard deviation per entry.

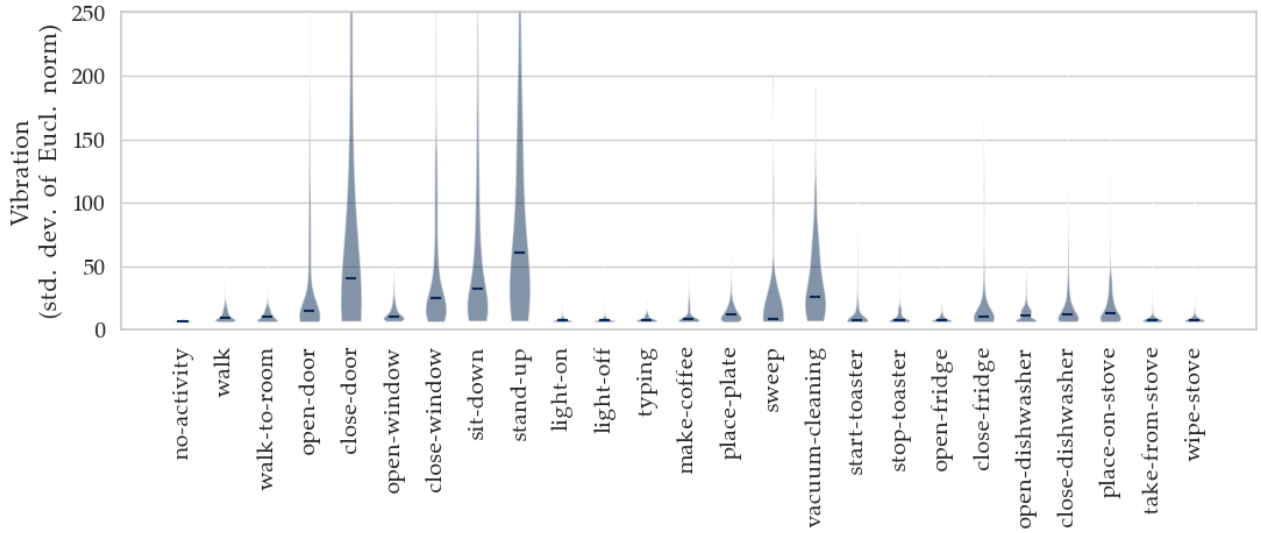

Figure A2. Variation of vibration data for each activity class. For each entry the standard deviation of the Euclidian norms was calculated.

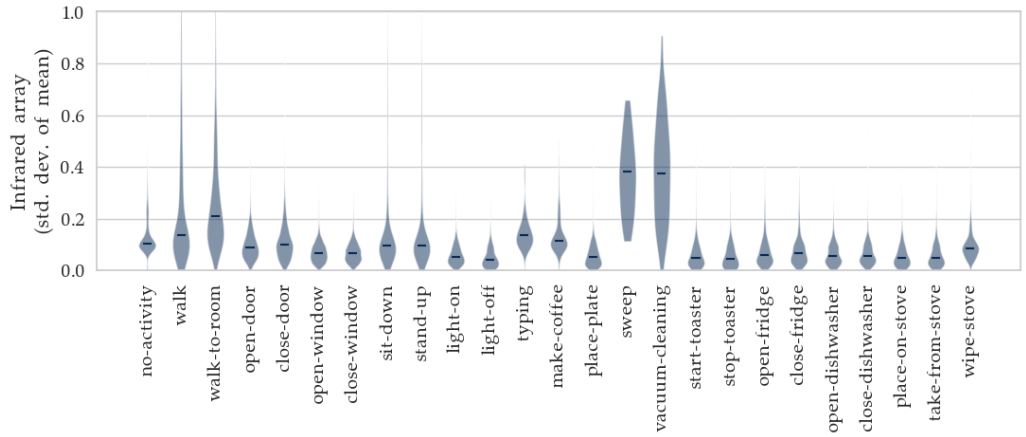

Figure A3. Variation of infrared array data for each activity class. For each entry, the standard deviation of the means was calculated.

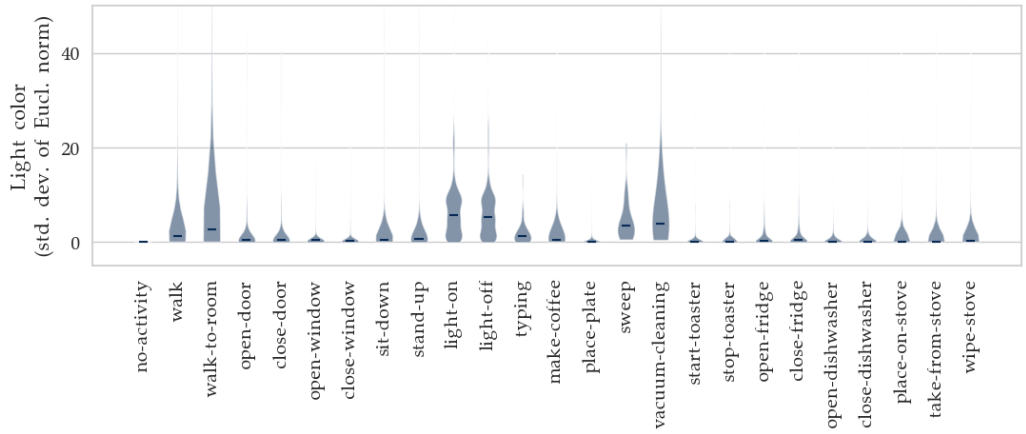

Figure A4. Variation of light color data for each activity class. For each entry, the standard deviation of the Euclidian norms was calculated.

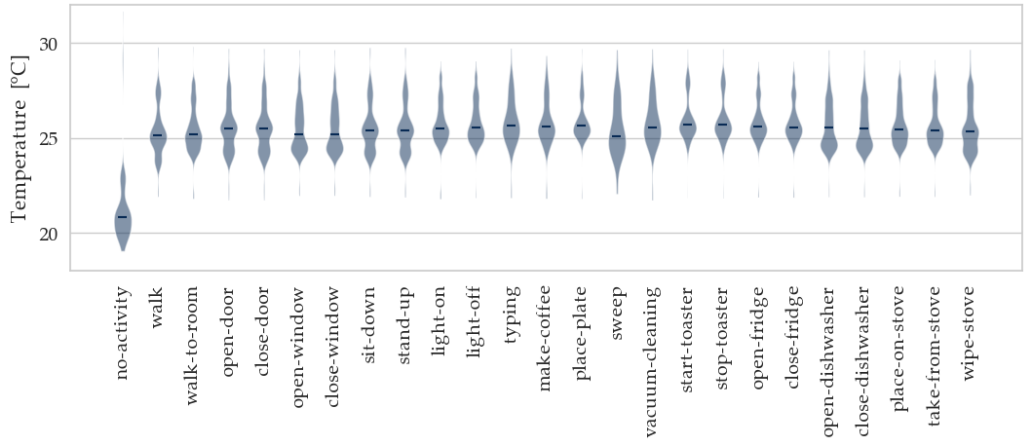

Figure A5. Temperature distribution for each activity class.

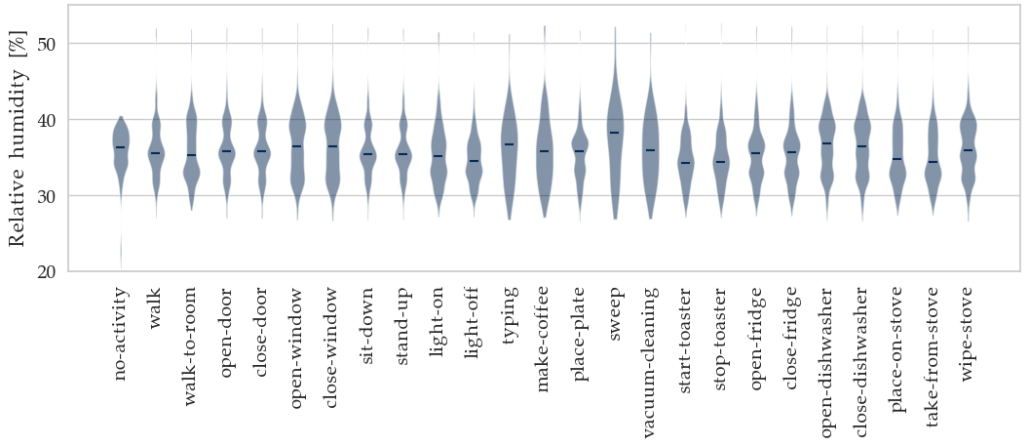

Figure A6. Humidity distribution for each activity class.

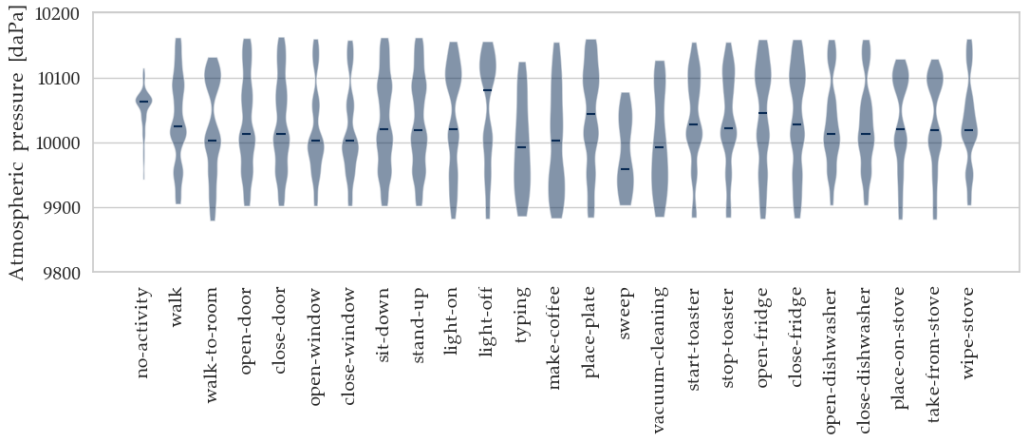

Figure A7. Pressure distribution for each activity class.

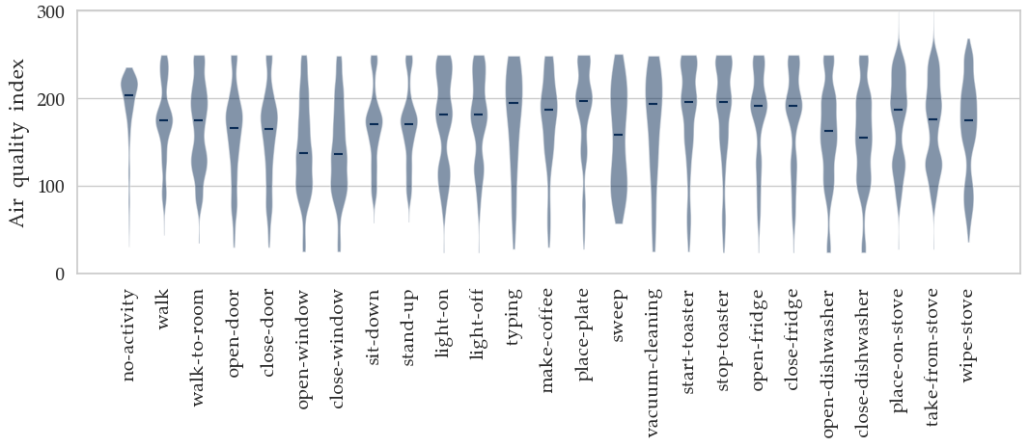

Figure A8. Air quality index distribution for each activity class.

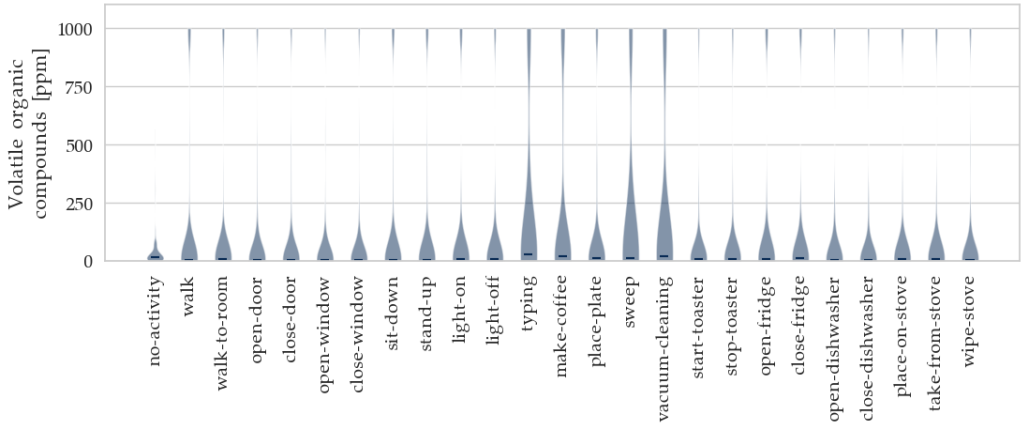

Figure A9. VOC distribution for each activity class.

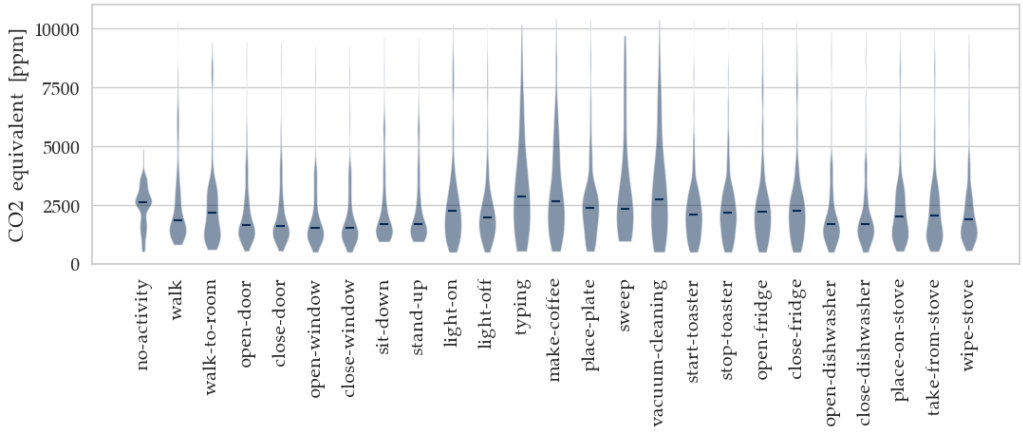

Figure A10. CO_2 equivalent distribution for each activity class.

For this analysis a single value was required per entry. For recordings with multiple measurements per entry, a single representative value was calculated for each entry. This was carried out as follows:

- Audio: the standard deviation of each entry.
- Vibration: the standard deviation of the Euclidean norms of the measurements (x, y and z component) in an entry.
- Infrared array: the standard deviation of the means of the measurement (8×8 matrix) in an entry.
- Light color: the standard deviation of the Euclidean norms of the measurements (red, green, blue and clear component) in an entry.
- As the environmental measurements contain only a single value, no transformation was necessary.

Note

[1] https://www.insta.de (accessed on 4 December 2024).

References

1. Bouchabou, D.; Nguyen, S.M.; Lohr, C.; LeDuc, B.; Kanellos, I. A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning. *Sensors* **2021**, *21*, 6037. [CrossRef] [PubMed]
2. Alam, G.; McChesney, I.; Nicholl, P.; Rafferty, J. Open Datasets in Human Activity Recognition Research—Issues and Challenges: A Review. *IEEE Sens. J.* **2023**, *23*, 26952–26980. [CrossRef]
3. Shoaib, M.; Bosch, S.; Incel, O.; Scholten, H.; Havinga, P. Fusion of Smartphone Motion Sensors for Physical Activity Recognition. *Sensors* **2014**, *14*, 10146–10176. [CrossRef] [PubMed]

4. Weiss, G. WISDM Smartphone and Smartwatch Activity and Biometrics Dataset. *UCI Mach. Learn. Repos.* **2019**, *7*, 133190–133202. [CrossRef]
5. Garcia-Gonzalez, D.; Rivero, D.; Fernandez-Blanco, E.; Luaces, M.R. A Public Domain Dataset for Real-Life Human Activity Recognition Using Smartphone Sensors. *Sensors* **2020**, *20*, 2200. [CrossRef] [PubMed]
6. Climent-Pérez, P.; Muñoz-Antón, Á.M.; Poli, A.; Spinsante, S.; Florez-Revuelta, F. Dataset of acceleration signals recorded while performing activities of daily living. *Data Brief* **2022**, *41*, 107896. [CrossRef] [PubMed]
7. Matey-Sanz, M.; Casteleyn, S.; Granell, C. Dataset of inertial measurements of smartphones and smartwatches for human activity recognition. *Data Brief* **2023**, *51*, 109809. [CrossRef] [PubMed]
8. Xu, L.; Wu, Q.; Pan, L.; Meng, F.; Li, H.; He, C.; Wang, H.; Cheng, S.; Dai, Y. Towards Continual Egocentric Activity Recognition: A Multi-Modal Egocentric Activity Dataset for Continual Learning. *IEEE Trans. Multimed.* **2024**, *26*, 2430–2443. [CrossRef]
9. Daniel Roggen, A.C. Opportunity Activity Recognition. *UCI Mach. Learn. Repos.* **2010**. [CrossRef]
10. Narayanan, R.M.; Zenaldin, M. Radar micro-Doppler signatures of various human activities. *IET Radar Sonar Navig.* **2015**, *9*, 1205–1215. [CrossRef]
11. Alsaify, B.A.; Almazari, M.M.; Alazrai, R.; Daoud, M.I. A dataset for Wi-Fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments. *Data Brief* **2020**, *33*, 106534. [CrossRef] [PubMed]
12. Alazrai, R.; Awad, A.; Alsaify, B.; Hababeh, M.; Daoud, M.I. A dataset for Wi-Fi-based human-to-human interaction recognition. *Data Brief* **2020**, *31*, 105668. [CrossRef] [PubMed]
13. Stork, J.A.; Spinello, L.; Silva, J.; Arras, K.O. Audio-based human activity recognition using Non-Markovian Ensemble Voting. In Proceedings of the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 509–514. [CrossRef]
14. Siantikos, G.; Giannakopoulos, T.; Konstantopoulos, S. Monitoring Activities of Daily Living Using Audio Analysis and a RaspberryPI: A Use Case on Bathroom Activity Monitoring. In *Information and Communication Technologies for Ageing Well and e-Health*; Röcker, C., O'Donoghue, J., Ziefle, M., Helfert, M., Molloy, W., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 736, pp. 20–32. [CrossRef]
15. Madhuranga, D.; Madushan, R.; Siriwardane, C.; Gunasekera, K. Real-time multimodal ADL recognition using convolution neural networks. *Vis. Comput.* **2021**, *37*, 1263–1276. [CrossRef]
16. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM Sigkdd Explor. Newsl.* **2011**, *12*, 74–82. [CrossRef]
17. Pfitzinger, T.; Wöhrle, H. Embedded Real-Time Human Activity Recognition on an ESP32-S3 Microcontroller Using Ambient Audio Data. In Proceedings of the 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 7–9 September 2023; pp. 459–464. [CrossRef]
18. Bosch. BME680 Datasheet. 2024. Available online: https://www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bme680-ds001.pdf (accessed on 20 November 2024).
19. Lavechin, M.; Métais, M.; Titeux, H.; Boissonnet, A.; Copet, J.; Rivière, M.; Bergelson, E.; Cristia, A.; Dupoux, E.; Bredin, H. Brouhaha: Multi-task training for voice activity detection, speech-to-noise ratio, and C50 room acoustics estimation. *arXiv* **2022**, arXiv:2210.13248. [CrossRef]
20. Bredin, H.; Yin, R.; Coria, J.M.; Gelly, G.; Korshunov, P.; Lavechin, M.; Fustes, D.; Titeux, H.; Bouaziz, W.; Gill, M.P. pyannote.audio: Neural building blocks for speaker diarization. *arXiv* **2019**, arXiv:1911.01255. [CrossRef]