Chapter 7
Applications for Environmental Sensing in EveryAware

Martin Atzmueller[1], Martin Becker[2,3], Andrea Molino[4], Juergen Mueller[1,3], Jan Peters[5], and Alina Sîrbu[6]

Abstract This chapter provides a technical description of the EveryAware applications for air quality and noise monitoring. Specifically, we introduce AirProbe, for measuring air quality, and WideNoise Plus for estimating environmental noise. We also include an overview on hardware components and smartphone-based measurement technology, and we present the according web backend, e. g., providing for real-time tracking, data storage, analysis and visualizations.

7.1 Introduction

Participatory sensing allows to approach many research questions. Such areas include understanding patterns, semantics, and dynamics of social behavior (e. g.,. [4, 7, 9, 11, 28, 29]) and its interaction with the sensor data collected by the corresponding applications. In this chapter, we introduce two such applications developed in the EveryAware project for collecting different environmental sensor information, specifically concerning air quality and noise pollution. For details on participatory environmental sensing please refer to Chapter "Sensing the environment" by Theunis, Stevens and Botteldooren in this part of the book. In order to facilitate the connection between sensor data and subjective data, both applications provide functionality to collect impressions, perceptions, or user defined contents in form of tags. Corresponding results are for example detailed in this book in Part III, Chapter "Emergence of awareness and behavioral changes: the EA lesson" by Gravino et al.

In particular, in this chapter we describe the AirProbe and WideNoise Plus applications. Both are utilizing smartphone-based data collection modules providing means for explicit subjective feedback and are backed by the versatile and flexi-

1. University of Kassel, Research Center for Information System Design · 2. University of Wuerzburg, Data Mining and Information Retrieval Group · 3. University of Hannover, L3S Research Center · 4. CSP · 6. VITO · 6. University of Bologna, Department of Computer Science and Engineering, Bologna, Italy

This is a preprint of a Chapter in: Loreto et al. (eds.) Participatory Sensing, Opinions and Collective Awareness. Springer Verlag, Berlin/Heidelberg, Germany. In Press

ble EveryAware backend, built upon the Ubicon platform. The latter is further introduced by Atzmueller, Becker and Mueller in Chapter "Collective Sensing Platforms" in this part of the book. For recording the data, AirProbe utilizes a special sensorbox for measuring air quality (based on data such as NO_2, CO, O_3, VOC, temperature, and humidity) which then transmits its data using a smartphone, enabling mobile data collection, cf., [19]. Similarly, WideNoise Plus provides the functionality for freely measuring environmental noise by using the smartphone's built-in microphone.

Both applications enable data access and inspection on the EveryAware web frontend: The smartphone gathers data and transmits it to our server where it is augmented and aggregated in order to provide comprehensive visualizations including for example noise pollution maps as mentioned in this book by Beate Weniger in Part II, Chapter "Cartographic Visualization of Noise and Aspects of Public Understanding of this Information". The applications have been used in several case studies, for providing first insights into collectively organized data collection in such objective and subjective contexts, cf., [3, 4, 9, 11, 29].

The remainder of this chapter is organized as follows: Section 7.2 describes the AirProbe application including the developed sensorbox and calibration steps. After that, Section 7.3 presents the WideNoise Plus application. Finally, Section 7.4 concludes with a summary and interesting options for future work. Contents of this chapter have been partially compiled from existing material, cf., [11, 29], in particular sections 7.2 and 7.3.

7.2 AirProbe

AirProbe is a mobile application for collectively monitoring air quality. To this end a calibrated low-cost sensor box has been developed: It displays the collected data on connected smartphones which are then used to upload the data to a central server. The data is then processed, enhanced and analyzed in order to generate feedback in form of statistics and map views displayed by a specialized frontend on the EveryAware web frontend. The following sections cover the basic components of the AirProbe stack, i.e. the sensor box, the smartphone application and the web frontend. Further details, e. g., about the data models used in the EveryAware web backend and the underlying Ubicon platform are described by Atzmueller, Becker and Mueller in Chapter "Collective Sensing Platforms" in this part of the book.

7.2.1 The EveryAware Sensor Box

Below, we first provide an overview on the EveryAware sensor box. After that, we give a technical description, before we discuss calibration steps and model learning.

7.2.1.1 Overview

The sensor box (see Figure 7.2) contains a sensor array of 8 commercially available gas sensors and two meteorological sensors (temperature and humidity). The gas sensor array consists of low-cost continuous sensors of CO, NOx, O3 and VOC, which are important pollutants in the urban outdoor environments. These pollutants are either directly emitted by vehicles or other combustion processes, or formed from emitted precursors in the vehicle exhaust. The gas sensors were examined by a range of performance tests under laboratory and outdoor conditions. Laboratory test cycles with sensor exposure to known gas concentrations were performed, whereas outdoor tests included comparison tests with reference measurements made with high-end reference monitors.

Fig. 7.1 The Arduino board of the sensor box.

7.2.1.2 Technical description

The sensor box electronic system has been designed with the purpose of being a low-cost, open and scalable platform. It includes basic storage (micro SD card), positioning (GPS) and communication (Bluetooth) capabilities, and accomodates a sensor shield able to host all gas sensors.

The architecture of the SensorBox is based on Arduino – an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software. The micro-controller on the board can be easily programmed to accomplish the tasks of the project. This system was chosen because of the simplicity with which it is possible to connect different component (shields) like GPS, Bluetooth or many others. Indeed, different shields are available on the market: they are ready to be connected and often there is a library to start programming. Considering that there is no need for complex data elaborations and that power consumption is an issue, Arduino is a good choice to have a prototype in short time. Furthermore, being an open-source project it is be possible, as we did, to review all the schematic and make a custom board with reduced dimension and cost, improving some parts if needed.

The development of the SensorBox followed various steps: after having tested each single device with the Arduino board, a first version of the SensorBox was produced in order to test the integration of the whole system and start testing. Then, a second version was designed with improvements on cost, weight, dimensions and signal integrity. The main step achieved in the second version is the implementation of a new electronic design based on a four layers Printed Circuit Board (PCB). The sensor boards are positioned in an air-tight housing. A continuous air flow is generated by a suction fen and an air outlet hole at the opposite side of the sensor box.

The design is based on Arduino components and it is completely open, so that anyone can reproduce and modify the hardware or even using the original hardware and develop different software to be run on it.

7.2.1.3 Sensor Box Calibration

Issues identified by laboratory and field testing included sensor sensitivity to temperature and humidity, sensor drift in time and sensitivity to other gasses. Additionally, measurement ranges were observed to vary between sensor boxes, with values difficult to map directly to pollutant concentrations. Hence one needs to calibrate devices against a reference in order to control for these issues and obtain a measurement meaningful for the user. Calibration is a mandatory step when using low cost or adapted sensors (see also chapters by J.Theunis, M. Stevens and D. Botteldooren and by Ferreira, Kostakos and Schweizer in this volume). The target pollutant selected in this study was black carbon (BC), motivated by several reasons:

- BC is a relevant pollutant in urban environment by its adverse health effects [20];
- BC is correlated with the gases that are measured by the sensor box, as learnt from the outdoor tests (Table 7.1);
- the availability of portable BC measurement devices (micro-aethalometers, AethLabs, Figure 7.2, also described in chapter by J.Theunis, M. Stevens and D. Botteldooren in this volume) which makes it possible to collect mobile BC data.

Table 7.1 Correlation between reference gas measurements at urban environment.

	Reference monitors				
	CO	NO	NO_2	O_3	BC
CO	1.00	0.77	0.62	-0.55	*0.83*
NO	0.77	1.00	0.76	-0.51	*0.89*
NO_2	0.62	0.76	1.00	-0.53	*0.81*
O_3	-0.55	-0.51	-0.53	1.00	*-0.54*
BC	0.83	0.89	0.81	-0.54	*1.00*

Calibration consisted in simultaneous measurements with the sensor boxes and the reference device (field calibration), and then training a model that is able to map the values measured by our sensor array with the values recorded by the reference. We have used artificial neural networks (ANNs) [27] for this regression task.

The micro-aethalometers provide high quality measurements of black carbon (BC), however at a much higher cost (about 30 times more expensive than our sensor box). The field-calibration has been inspired by the works of Carotta et al. [14] [15] [13], Tsujita et al. [30], Kamionka et al. [23] and De Vito et al. [16] [17].

Fig. 7.2 Microaethalometer: device used as a reference for calibration.

Three types of data were used to train a calibration model, to account for three possible use cases. These included stationary data where all sensor boxes were collocated (in the same place), mobile measurements performed with one or two boxes at a time, and indoor data. Calibration models were trained in four different cities: Antwerp, London, Kassel and Turin. The representativity of the datasets is crucial. We increased the representativity of the data by placing the measurement equipment within the final area of deployment (city-specific), by collecting data continuously during day and nighttime (stationary data) and by collecting data over quite a long period (10 days) right before the final deployment (sensor boxes were used in the final test case two weeks later).

7.2.1.4 Data Preprocessing

Although initially the possibility of building one calibration model for each box was intended, this would not have scaled very well, so we explored the possibility of building one model for all sensor boxes. This has also the advantage that data from multiple boxes can be used, resulting, possibly, in better modelling performance. For this, we first compared sensor output for the different sensor boxes. Gas sensors produce a signal with a value in $[0,5]$, however only part of this interval is actually used. It is the setting of the potentiometer (done by hand by our engineers) that determines the exact range, so differences between the output of the same sensors on different sensor boxes are impossible to avoid. However, in principle, the fluctuations should correlate. Figure 7.3 displays an example of two collocated sensor boxes and the sensor responses. While some sensors have similar ranges, some others do not (for instance the dashed red sensor). However it appears that a linear scaling could bring the values in the same interval.

Considering this, a scaling procedure was employed in order to enable the use of one model for all boxes. Using the entire stationary data when all boxes were performing measurements together, the active range of each sensor was determined and then all data rescaled so that the active range falls to interval $[0.2, 0.7]$. This allows for measurements outside the range shown during the stationary measure-

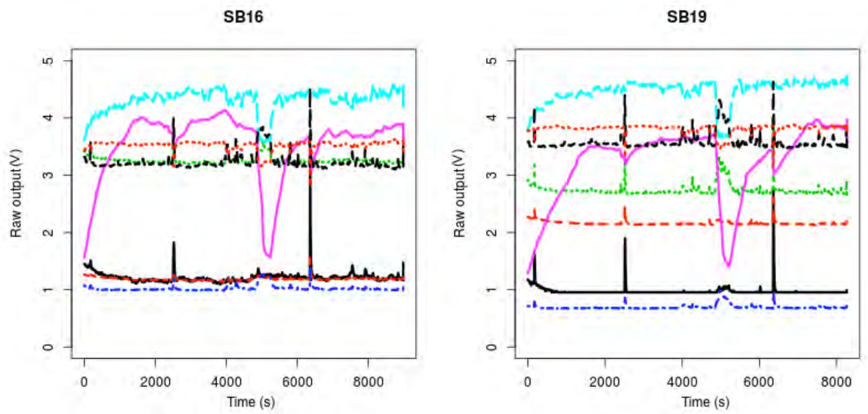


Fig. 7.3 Comparison of sensor output for two sensor boxes.

ments to appear in the future. Any future data will be scaled using the same scaling parameters extracted from the stationary collocated data. Temperature and humidity, that also vary from one box to another (due to small differences in air flow or position) were scaled to interval $[0.1, 0.4]$. This procedure allowed us to obtain an unique model for all sensor boxes in each location, instead of individual ones for each box. Tests indicated no significant differences in model performance using sensor box specific models or a more general model that is applied to sensor box specific rescaled sensor box signals.

A different issue was data variability, both in BC values and sensor response. The BC values were post-processed by a noise reduction algorithm [22] to lower the high-frequency instrument noise that is observed when measuring at high frequency. BC levels were further smoothed by averaging over a 5 minute moving window. This value was deemed suitable by comparing outputs from two aethalometers, which become highly correlated at this resolution. So the BC value obtained from the model represents an average over the last five minutes of exposure. Comparing the time series, we observed a lower sensitivity of the sensor box compared to the aethalometer, leading to delays in sensor response. To account for this, we used a smaller time window (60 seconds) to smooth the sensor output.

7.2.1.5 Model training and testing

After preprocessing, training and testing datasets were obtained for each location by combining all data types available. An ANN model for each location was obtained from the training data using backpropagation. Empirical tests showed that best performance was obtained with an ANN with one hidden layer of 10 neurons, as shown in Figure 7.4.

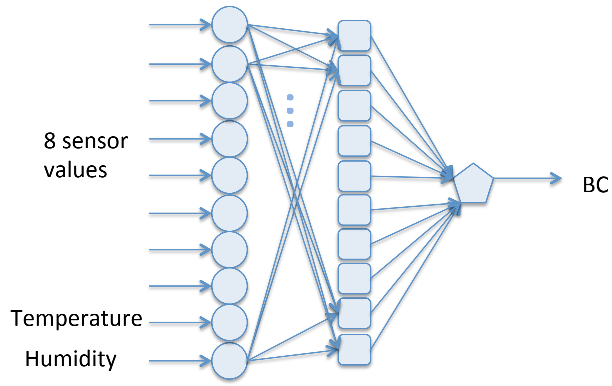


Fig. 7.4 ANN topology for our calibration problem.

All training has been repeated several times and the model with best behaviour on the test data was selected. Model performance was evaluated using 3 criteria: R^2, root mean squared error (RMSE) and Pearson correlation coefficient between the modelled and the measured BC time series.

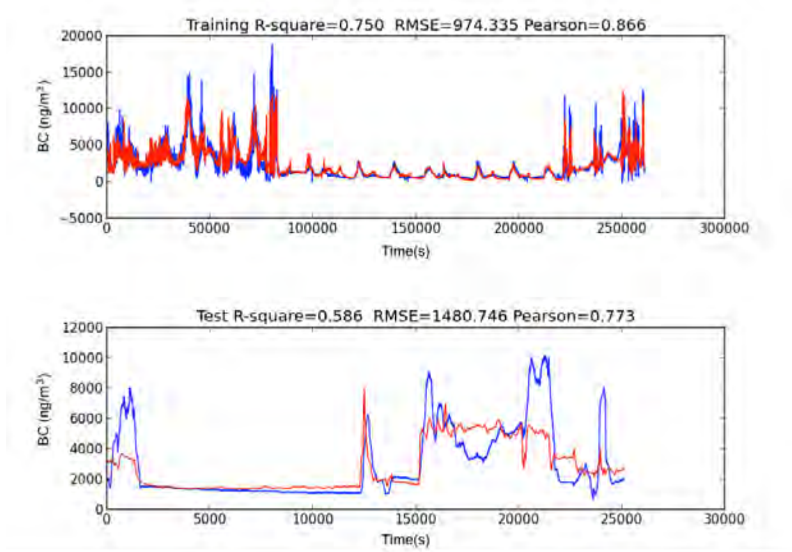


Fig. 7.5 Model performance in Turin

Figure 7.5 shows the result of calibration for Turin, obtained before the final test case. Two time series are shown, one in red, representing the model output, one in blue representing the data measured by the aethalometer, with evaluation criteria shown at the top, indicating very good overall results. While on training data the two time series match very well, on test data the model appears to miss some of the fluctuations seen in the BC values. This shows that, in general, the model is successful in identifying general trends in the pollution levels. However, sharp and short peaks are not handled very well by the model, and this is due to the lower sensitivity of the low cost sensors and their delayed response. However, the performance obtained was enough for the purposes of the project, i.e. participatory mapping of pollution with multiple devices, for enhancing environmental awareness.

The ANN model was implemented both in the AirProbe application, to give the user real time feedback from the sensor box, but also server side. This approach was taken due to the fact that the sensor box has two working modes, one online and one offline. Computing model output for all offline records would have been too computationally expensive for an average smartphone, while server side this was not an issue.

7.2.2 AirProbe Smartphone Application

AirProbe[1] is a smartphone application for Android. It is used to read the data from the AirProbe sensor box, allows users to view, browse and annotate the data, and to upload it to the EveryAware server. Then, the data is further analyzed and processed in order to provide additional statistics and views via the EveryAware web platform.

To associate the user with the uploaded data, the user first registers her account from the EveryAware backend within the AirProbe smartphone application. Afterwards there are three operational modes the AirProbe smartphone application provides: the Live Track mode, which allows to view the currently measured air quality, the Browsing mode, which allows to browse and view already collected data, and the Synchronization mode, which is used for actively managing the uploading process. In the following, we will describe these three modes in detail.

7.2.2.1 Live Track mode

The Live Track mode allows to monitor air quality components in real time. When starting this mode, first the application will search for Bluetooth devices nearby and present the user with a list of found devices. Once the user has selected the sensor box, AirProbe starts displaying real time data collected by the sensor box, using the Bluetooth connection. The interface of the Live Track mode is composed of three different views accessible from their corresponding tabs (Figure 7.6):

---

[1] The AirProbe application is freely available for the Android platform and can be installed from Google PlayStore.

Map, where users can follow their own live track. The track is represented with different colours, depending on real-time black carbon levels. The user can also add annotations and share them on social networks (Facebook/Twitter), using the buttons at the top right corner. The track length to be shown on the map can be of 5, 15, 60 minutes. Live updating of the current position can be switched on and off, through the top left buttons. The bar at the top represents the black carbon value using a coloured scale (from a blue/low value to a brown/high value).

Graph, where the user can see a graph of the black carbon measurements as well as of the raw data from pollutant sensors, in a variable time interval ranging from 1 to 30 minutes. The user can query the value registered by each sensor by tapping on the series. The graph is updated every two seconds.

Monitor, where users can access statistics about collected data, connection information, the status of the sensor box and the installed sensors.

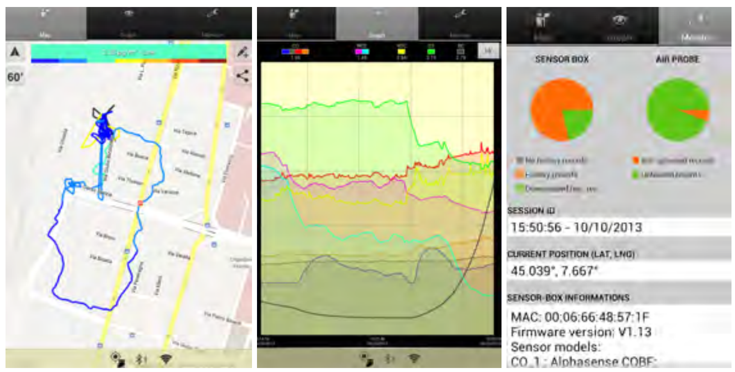


Fig. 7.6 AirProbe screenshots: Live mode. AirProbe uses the Google Maps API to display maps (©2014 Google [21]).

7.2.2.2 Browsing mode

The Browsing mode enables the user to access aggregated air quality measurements in the area from the server and browse their own tracks which are still on the phone. This working mode does not require an active Bluetooth connection to a sensor box. It is composed by three views, accessible from their corresponding tabs:

Map, where the user can see the black carbon levels around his current position (Figure 7.7), by pressing the "Get nearby BC levels" button. If a track from "My-Tracks" tab is selected, it is displayed on the map. The global black carbon levels and selected track can be shown together.

Graph, where the raw pollutant and black carbon evolution, calculated for a selected track, are shown. However, only tracks which have been recorded in Live Mode have black carbon data.

My Track, where the list of tracks available on the mobile device is shown. Older tracks are automatically deleted once they have been uploaded to the server and a configurable time interval since their creation has passed.

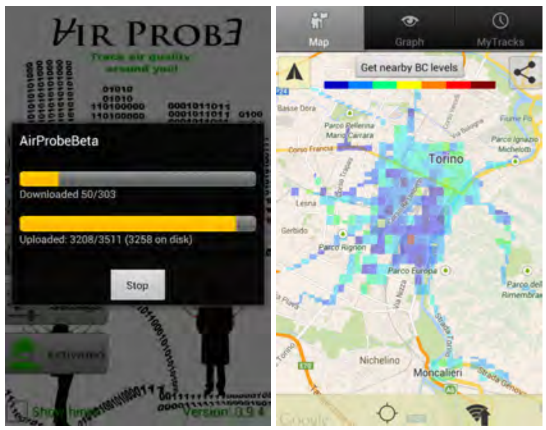


Fig. 7.7 AirProbe screenshots: Synchronization and Black Carbon map. AirProbe uses the Google Maps API to display maps (©2014 Google [21]).

7.2.2.3 Synchronization mode

In this working mode, AirProbe reads data from the sensor box and uploads them to the EveryAware server (Figure 7.7). This allows the box to run without a smartphone. The user can then send the data to server in suitable conditions (e.g. where battery lifetime and/or connection billing are not a problem).

7.2.3 AirProbe Web Application

The AirProbe web application is part of the EveryAware backend [12] embedded into the Ubicon platform [3, 4] as described by Atzmueller, Becker and Mueller in Chapter "Collective Sensing Platforms" in this part of the book. It processes the data it receives from the AirProbe smartphone app, see Section 7.2.2, cleans it, applies Black Carbon calculation as described in Section 7.2.1.3, and provides several statistics and views for the user to analyze and understand her data. Furthermore, it supports case studies like the "AirProbe International Challenge" which is further described by Sîrbu et al. [29] and in Chapter "Experimental assessment of the emergence of awareness and its influence on behavioral changes: the EveryAware lesson" by Gravino et al. in Part III of this book. In the following, we briefly introduce some statistics and views the web application provides and summarize the functionality used for the mentioned case study.

7.2.4 Statistics and Visualizations

For the AirProbe module the visualized information is represented by several views of the data including a map with different information layers as well as several global and personal statistics. The OpenStreetMap-based[2] map view visualizes the collected data on a map which allows for an easy access to the data as well as for obtaining first insights. It provides a quantitative view by aggregating samples using clusters, grids, as well as a heatmap view in order to emphasize the covered area on a global and on a personal level (see Figure 7.8).

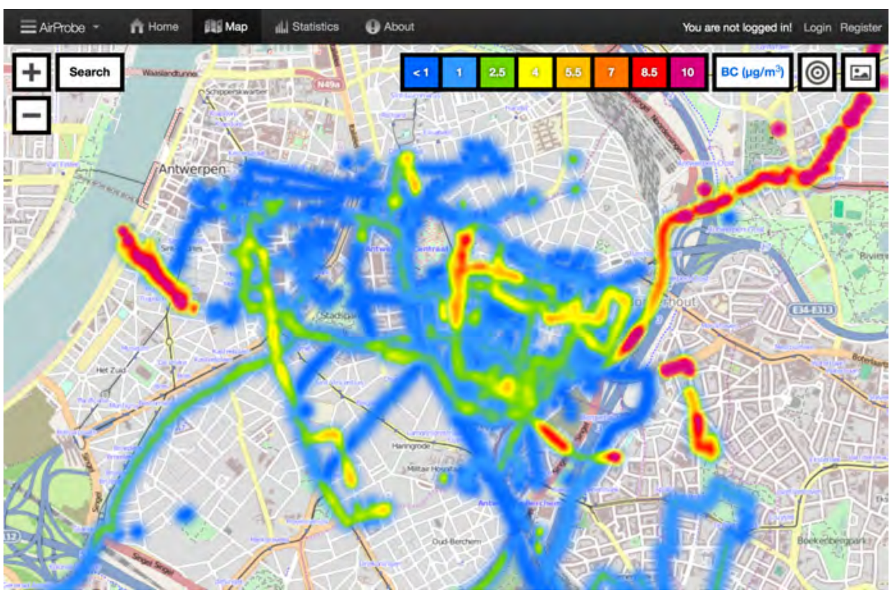
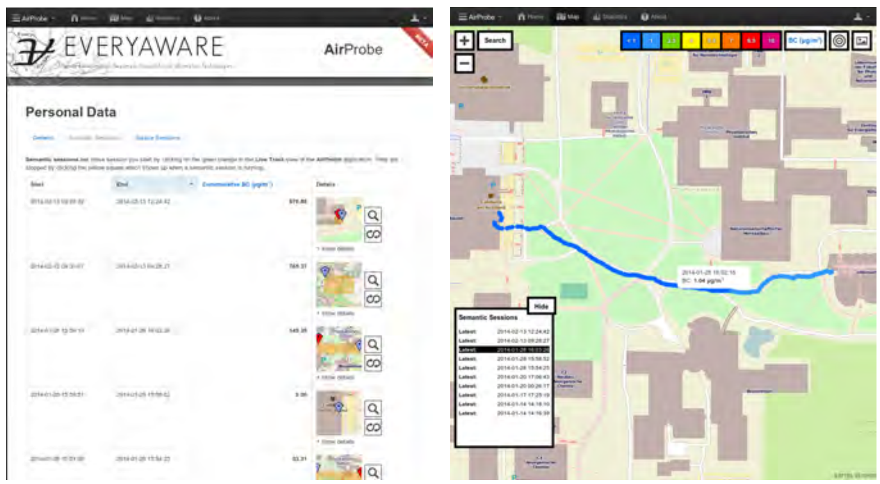


Fig. 7.8 A screenshot of a heatmap on the *map page* of AirProbe. The website map and heatmap were generated using in-house developed tools and OpenStreetMap data (©*OpenStreetMap contributors* for map data, used and redistributed under the CC-BY-SA licence [1]) .

[2] `http://openstreetmap.org/`

Further statistics calculated by the AirProbe application include summaries like latest overall measurement activity or air quality averages. Also, personal user profiles are available. Among other things, these profiles list “measurement sessions”. For each session a short summary of the user’s measuring activities is given. Sessions can further be viewed and explored for example by replaying the measuring process. A personal sessions overview can be seen in Figure 9(a). A view for exploring personal sessions can be seen in Figure 9(b).



(a) This AirProbe view shows a user’s personal sessions.

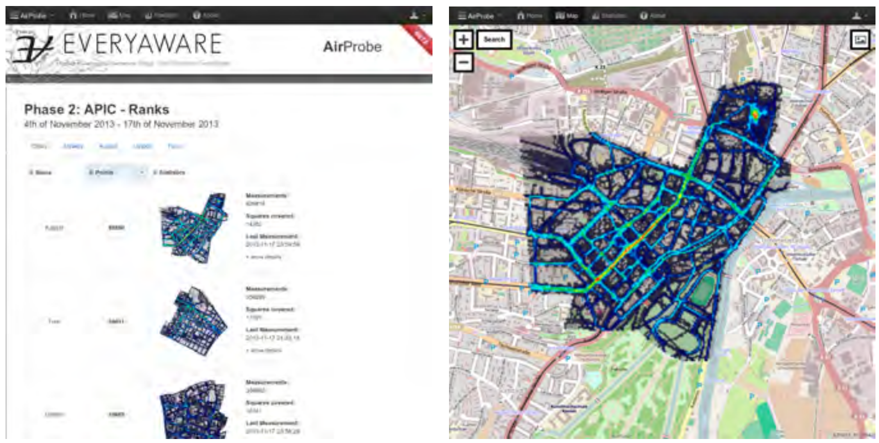
(b) This AirProbe view shows a view for exploring individual user sessions.

Fig. 7.9 AirProbe personal measurement sessions visualizations. The website map and track visualisation were generated using in-house developed tools and OpenStreetMap data (©*OpenStreetMap contributors* for map data, used and redistributed under the CC-BY-SA licence [1]) .

7.2.5 APIC Rankings

In addition to the global and personal statistics, the web interface provides feedback for the users participating in case studies like the APIC (‘AirProbe International Challenge”) [29]. The APIC case study was held in order to gather large amounts of air quality samples and behavioral shift patterns using the sensorboxes in the four cities Antwerp, Kassel, London, and Turin.

In order to keep the motivation and competitiveness as high as possible for the teams playing, we implemented a ranking mechanism balancing repetitive sampling and coverage. The map was divided into 10 by 10 meter grids. One point was given to a team when sampling within one such grid cell. When a team received a point in a particular cell, the player did not receive a point from this grid cell for half an hour. The results for each city as well as for each team have been visualized and updated in regular intervals on the AirProbe website as can be seen in Figure 7.10. Figure 10(a) shows the ranking of each city visualizing the coverage and providing several statistics. Figure 10(b) shows a detailed view of the point-coverage of the city.



(a) APIC city ranking. (b) APIC point coverage for Kassel.

Fig. 7.10 APIC ranking visualizations. The website map and heatmaps were generated using in-house developed tools and OpenStreetMap data (©*OpenStreetMap contributors* for map data, used and redistributed under the CC-BY-SA licence [1]) .

7.3 WideNoise Plus

There are various kinds of pollution that get often on the first page of newspapers. However, *noise* pollution is rarely cited even though it is something that constantly surrounds us even if we are not aware.

WideNoise Plus has been developed to record, monitor, and analyze such noise pollution and helps to better understand the user's soundscape. Its predecessor was developed by WideTag Inc. and was acquired and extended by the EveryAware team. Good and bad noise is not the same as loud and silent noise. One vivid examples is a rock concert. It is extremely loud on the one hand, but is mainly for pleasure on the other hand. Thus, WideNoise Plus was extended in order to support subjective annotations in order to reflect the perceived quality of the recorded noise.

WideNoise Plus is running for more than three years now. It is used, e.g., by the citizens around the Heathrow airport to monitor noise pollution caused by air traffic. Until now we collected more than 54,700 noise samples recorded by over

16,800 devices from all over the world. Insights into the corresponding data are reported, for example, by Becker et al. [11], Atzmueller et al. [9], or in this book in Part III, Chapter "Emergence of awareness and behavioral changes: the EveryAware lesson" by Gravino et al..

As a related system, Kanjo [24] presented the first system for collecting noise data with mobile phones and discusses its implementation on a technical level. There are several existing platforms dedicated to specialized sensor data types. Maisonneuve et al. [26] present an approach for monitoring the noise pollution by the general public using the NoiseTube[3] system. AirCasting[4] is another platform, which allows users to upload information about surrounding noise and air quality using their mobile phones. In contrast to these systems, WideNoise Plus focuses on user feedback in addition to recording noise. This feedback comes in two forms: user estimates of noise, and subjective data. User estimates help the user to gauge if they asses the noise around them correctly. At the same time subjective data is collected in the form of perceptions or tags. Perceptions range for example from a "social" feeling to solitude, or from "love" to "hate", and tags may include anything the user feels is relevant about the recorded noise. Thus, when accessing statistics the user can find patterns in how noise affects her. Overall, and in contrast to other systems, WideNoise Plus is more focused on feedback from the user, thus, allowing for a enhanced learning process with regard to awareness concerning noise pollution.

The remainder of this section is based on the article by Becker et al. [11].

7.3.1 Smartphone Application

WideNoise Plus was developed for the two major mobile systems iOS and Android in order give access to as many people as possible. It records the noise level of the soundscape using the build-in microphone of the smartphone. No audio track is created during recording, only the loudness level every 0.5 seconds; therefore, the privacy of the user is ensured. The anonymous noise levels are transmitted to our application server (see Section 7.3.2) through a RESTful web service. Both, sensor data and subjective perceptions are required to create a full sound report, so that the application consists of two main parts:

1. Objective noise recordings.
2. Subjective annotations.

The noise recording part gives users a tool to take a noise sample through the smartphone microphone. During the recording, the user is asked to guess the current noise level using a slider where a decibel scale is mapped. The user has also the possibility of extending the default sampling time of 5 seconds to 10 or 15 seconds. In this way, the app will perform a longer measurement while the user gets more time

[3] `http://noisetube.net/`

[4] `http://aircasting.org/`

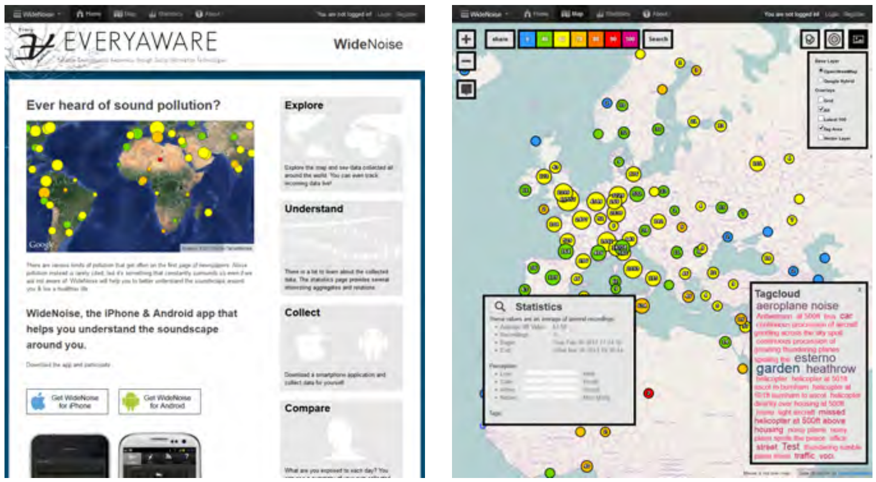
to make the guess. After the recording, the noise level expressed in decibels (dB) is shown and compared to the estimation of the user. The sound level is illustrated with an icon that allows a better understanding of the measured decibel value. The icon categorize the noise level into the following groups (see Figure 11(a)): falling feather (i.e., $[0,30]$ dB), sleeping cat (i.e., $]30,60]$ dB), TV show (i.e., $]60,70]$ dB), car engine (i.e., $]70,90]$ dB), dragster (i.e., $]90,100]$ dB), t-rex (i.e., $]100,115]$ dB), and rock concert (i.e., $]115,120]$ dB). After the recording view, the users are asked to express their own subjective impression about the recorded noise. At first, they can express their opinion by moving four different sliders associated to the following concepts (see Figure 11(b)): love/hate, calm/hectic, alone/social, and nature/man-made. At second, they can associate free text tags to the noise to further express their impression (see Figure 11(c)). Once the subjective information is attached, all the information collected by the application is sent to the web application server as soon as a working data connection is available. WideNoise Plus allows users to view a community map displaying the average noise level at nearby locations, by relying on the statistical elaboration provided by the server (see Figure 11(d)). As an integration with social networks, users can also share their own recordings via Twitter and Facebook.

(a) Recording screen (b) Perceptions screen (c) Tagging screen (d) Monitoring screen

Fig. 7.11 Screenshots from the WideNoise Plus Android application.

7.3.2 Web Application

The WideNoise Plus web application (see Figure 12(a)) is part of the EveryAware backend [12] embedded into the Ubicon platform [3,4]. It aggregates, summarizes, and illustrates noise related data collected by the smartphone application. It provides several statistics for global and personal levels and renders a map for spatial exploration (see Figure 12(b)). Additionally, the web application provides useful information about the smartphone application and its history.



(a) Main page (b) Map page

Fig. 7.12 Screenshots from the WideNoise Plus web application.

The web application provides several statistics on global and personal levels. These statistics help the user to explore and understand the data as well as to observe trends in usage patterns or noise distributions. The statistics include but are not limited to:

- The number of recordings on every day during the last two weeks.
- Contributing user activity distribution.
- The number of recordings during the last three days aggregated by continent.
- Relation between user estimates and actual decibel values.
- The three last recordings with their measured and guessed decibel value, a timestamp, the name of the location (e.g., Kassel, Germany) and the subjective annotations.
- User rankings including users with most samples, the most active users.
- The average noise level during the last day, month, and year.
- The number of registered users (those with a user account on the web server), linked devices (those registered users that linked their device to their account), and the overall number of devices.
- The average noise level per day during the last two weeks.
- A tag cloud for the last week, month, year, and for all collected data.

The web application also provides personalized content. Users can access their personal data and statistics via their personal page, e.g., for information on their own measuring behavior. The personal page also provides a KML export of the user's measurements as an alternative to the map visualization.

The map page is one of the most powerful features of the web application (see Figure 12(b)). For example, a cluster and a grid view are summarizing the noise data providing detailed information on demand. Averages of the measured noise and of the perceptions recorded by the smartphone application are available. For registered users a personalized view on the data is provided. Furthermore, a tag cloud characterizes the summarized data by its semantic context. To support social activities, the ability to forward the current view of the map to Twitter or Facebook was introduced. This allows the user to directly share and discuss interesting areas and sample distributions with friends or followers. Another feature of the map is the tracking of incoming measurements in real-time. Thus, the map connects the user to the ongoing measurement process all over the world.

7.4 Conclusions

For data collection, mobile applications (AirProbe and WideNoise) have been developed and designed to measure air quality and noise, respectively. At the same time these applications enable users to contribute subjective data. The Widenoise application uses the integrated microphone in smartphones to record noise. The AirProbe application makes use of a unique sensor box which has been designed using of-the-shelf sensors, hardware and data handling technology. Lab and outdoor experiments with the sensor box resulted in the development of a calibration model to estimate black carbon concentrations from the sensor measurements. The (mobile) measurements are then transferred to a web platform.

The design of web-based infrastructures has a great influence both on data quantity and quality, and hence also on the additional value which can be generated by analyzing the resulting datasets. Therefore, appropriate methods and techniques of acquiring and handling such data efficiently played a central role in the development of the presented applications, built on top of the Ubicon software platform [4], which enables the observation of physical and social activities.

For future work, enhancing integrated exploratory techniques, e. g., [2, 9] and extended visualization methods for geo-social data that also provide for detailed data introspection techniques, e. g., [7, 10] are promising options. This also concerns methods for integrated detection and analysis of anomalous and exceptional patterns, e. g., [5, 8]. Furthermore, integrating advanced processing features, based on techniques for handling large structured and unstructured data, e. g., [6, 25] also in the spirit of Big Data, e. g., [18, 31], seem further worthwhile directions to consider.

References

1. http://www.openstreetmap.org/copyright (Date accessed: 03/02/2015)
2. Atzmueller, M.: Subgroup Discovery – Advanced Review. WIREs: Data Mining and Knowledge Discovery **5**(1), 35–49 (2015). DOI 10.1002/widm.1144
3. Atzmueller, M., Becker, M., Doerfel, S., Kibanov, M., Hotho, A., Macek, B.E., Mitzlaff, F., Mueller, J., Scholz, C., Stumme, G.: Ubicon: Observing social and physical activities. In: IEEE International Conference on Cyber, Physical and Social Computing, CPSCom 2012, Besancon, France, 20-23 November, 2012, pp. 317–324. IEEE, Washington, DC, USA (2012). DOI 10.1109/GreenCom.2012.75
4. Atzmueller, M., Becker, M., Kibanov, M., Scholz, C., Doerfel, S., Hotho, A., Macek, B.E., Mitzlaff, F., Mueller, J., Stumme, G.: Ubicon and its Applications for Ubiquitous Social Computing. New Review of Hypermedia and Multimedia **20**(1), 53–77 (2014). DOI 10.1080/13614568.2013.873488
5. Atzmueller, M., Doerfel, S., Mitzlaff, F.: Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Information Sciences **329**, 965–984 (2016)
6. Atzmueller, M., Kluegl, P., Puppe, F.: Rule-based information extraction for structured data acquisition using textmarker. In: Lernen, Wissensentdeckung und Adaptivität, LWA 2008, Würzburg, Germany – October 06-08, 2008. Proceedings. University of Würzburg, Würzburg, Germany (2008)
7. Atzmueller, M., Lemmerich, F.: Exploratory Pattern Mining on Social Media using Geo-References and Social Tagging Information. IJWS **2**(1/2) (2013)
8. Atzmueller, M., Lemmerich, F., Krause, B., Hotho, A.: Who are the Spammers? Understandable Local Patterns for Concept Description. In: Proc. 7th Conference on Computer Methods and Systems (2009)
9. Atzmueller, M., Mueller, J., Becker, M.: Mining, Modeling and Recommending 'Things' in Social Media, chap. Exploratory Subgroup Analytics on Ubiquitous Data. No. 8940 in LNAI. Springer, Heidelberg, Germany (2015)
10. Atzmueller, M., Puppe, F.: A Case-Based Approach for Characterization and Analysis of Subgroup Patterns. Journal of Applied Intelligence **28**(3), 210–221 (2008)
11. Becker, M., Caminiti, S., Fiorella, D., Francis, L., Gravino, P., Haklay, M., Hotho, A., Loreto, V., Mueller, J., Ricchiuti, F., Servedio, V.D.P., Sîrbu, A., Tria, F.: Awareness and learning in participatory noise sensing. PLOS ONE **8**(12), e81,638 (2013). DOI 10.1371/journal.pone.0081638
12. Becker, M., Mueller, J., Hotho, A., Stumme, G.: A generic platform for ubiquitous and subjective data. In: 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing , UbiComp 2013; 1st International Workshop on Pervasive Urban Crowdsensing Architecture and Applications, PUCAA 2013, Zurich, Switzerland – September 8-12, 2013. Proceedings, pp. 1175–1182. ACM, New York, NY, USA (2013). DOI 10.1145/2494091.2499776
13. Carotta, M., Benetti, M., Ferrari, E., Giberti, A., Malagu, C., Nagliati, M., Vendemiati, B., Martinelli, G.: Basic interpretation of thick film gas sensors for atmospheric application. Sensors and Actuators B: Chemical **126**(2), 672 – 677 (2007)
14. Carotta, M., Martinelli, G., Crema, L., Gallana, M., Merli, M., Ghiotti, G., Traversa, E.: Array of thick film sensors for atmospheric pollutant monitoring. Sensors and Actuators B: Chemical **68**(1 - 3), 1 – 8 (2000)
15. Carotta, M., Martinelli, G., Crema, L., Malagu, C., Merli, M., Ghiotti, G., Traversa, E.: Nanostructured thick-film gas sensors for atmospheric pollutant monitoring: quantitative analysis on field tests. Sensors and Actuators B: Chemical **76**(1 - 3), 336 – 342 (2001)
16. De Vito, S., Massera, E., Piga, M., Martinotto, L., Di Francia, G.: On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators B: Chemical **129**(2), 750 – 757 (2008)
17. De Vito, S., Piga, M., Martinotto, L., Di Francia, G.: Co, no2 and nox urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. Sensors and Actuators B: Chemical **143**(1), 182 – 191 (2009)

18. Dean, J., Ghemawat, S.: Mapreduce: Simplified data processing on large clusters. Communications of the ACM **51**(1), 107–113 (2008). DOI 10.1145/1327452.1327492
19. Elen, B., Theunis, J., Ingarra, S., Molino, A., Van, J., den Bossche, Reggente, M., Loreto, V.: The everyaware sensorbox: A tool for community-based air quality monitoring. In: Sensing a Changing World Workshop, Wageningen, The Netherlands, 9 May, 2012. Proceedings (2012)
20. EPA: Report to congress on black carbon - external peer review draft. Tech. Rep. EPA-450/D-11-001, EPA (2010)
21. Google: http://www.google.com/permissions/geoguidelines/attr-guide.html (Date accessed: 03/02/2015)
22. Hagler, G., Yelverton, T., Vedantham, R., Hansen, A., Turner, J.: Post-processing method to reduce noise while preserving high time resolution in aethalometer real-time black carbon data. Aerosol and Air Quality Research **11**, 539 – 546 (2011)
23. Kamionka, M., Breuil, P., Pijolat, C.: Calibration of a multivariate gas sensing device for atmospheric pollution measurement. Sensors and Actuators B: Chemical **118**(1 - 2), 323 – 327 (2006)
24. Kanjo, E.: Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping. Mobile Networks and Applications **15**(4), 562–574 (2010). DOI 10.1007/s11036-009-0217-y
25. Kluegl, P., Atzmueller, M., Puppe, F.: Meta-level information extraction. In: KI 2009: Advances in Artificial Intelligence. 32nd Annual German Conference on AI, Paderborn, Germany, September 2009. Proceedings, *Lecture Notes in Computer Science*, vol. 5803, pp. 233–240. Springer, Berlin / Heidelberg, Germany (2009). DOI 10.1007/978-3-642-04617-9_30
26. Maisonneuve, N., Stevens, M., Ochab, B.: Participatory noise pollution monitoring using mobile phones. Information Polity **15**(1-2), 51–71 (2010). DOI 10.3233/IP-2010-0200
27. Mitchell, T.M.: Machine learning. 1997. Burr Ridge, IL: McGraw Hill **45** (1997)
28. Mitzlaff, F., Atzmueller, M., Stumme, G., Hotho, A.: Semantics of user interaction in social media. In: Complex Networks IV. Proceedings of the 4th Workshop on Complex Networks CompleNet 2013, *Studies in Computational Intelligence*, vol. 476, pp. 13–25. Springer, Heidelberg, Germany (2013). DOI 10.1007/978-3-642-36844-8_2
29. Sîrbu, A., Becker, M., Caminiti, S., De Baets, B., Elen, B., Francis, L., Gravino, P., Hotho, A., Ingarra, S., Loreto, V., Molino, A., Mueller, J., Peters, J., Ricchiuti, F., Saracino, F., Servedio, V.D.P., Stumme, G., Theunis, J., Tria, F., Van den Bossche, J.: Participatory Patterns in an International Air Quality Monitoring Initiative. arXiv **1503.07730** (2015)
30. Tsujita, W., Yoshino, A., Ishida, H., Moriizumi, T.: Gas sensor network for air-pollution monitoring. Sensors and Actuators B: Chemical **110**(2), 304 – 311 (2005)
31. Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I.: Spark: Cluster Computing with Working Sets. In: Proc. USENIX Conference on Hot Topics in Cloud Computing, HotCloud'10, pp. 10–10. USENIX Association, Berkeley, CA, USA (2010)