Experimental Assessment of the Emergence of Awareness and Its Influence on Behavioral Changes: The Everyaware Lesson

Chapter · May 2017

DOI: 10.1007/978-3-319-25658-0_16

Pietro Gravino
Sapienza University of Rome

Alina Sîrbu
Dublin City University

Vito Domenico Pietro Servedio
CSH Vienna

Vittorio Loreto
Sapienza University of Rome

Chapter 2
Experimental assessment of the emergence of awareness and its influence on behavioral changes: the EveryAware lesson

Pietro Gravino[1], **Alina Sîrbu**[2], **Martin Becker**[3,4], **Vito D. P. Servedio**[5,1] **and Vittorio Loreto**[1,6,7]

[1]Physics Department, Sapienza University, Rome, Italy

[2]University of Bologna, Department of Computer Science and Engineering, Bologna, Italy

[3]Department for Artificial Intelligence and Applied Computer Science, University of Würzburg, Würzburg, Germany

[4]L3S Research Center, Leibniz Universität, Hannover, Germany

[5]Institute for Complex Systems (ISC-CNR), Rome, Italy

[6]Complex Networks and Systems Lagrange Laboratory, Institute for Scientific Interchange Foundation, Turin, Italy

[7]SONY-CSL Computer Science Lab, Paris, France

[pietro.gravino@gmail.com]

2.1 Overview

The emergence of awareness is deeply connected to the process of learning. In fact, by learning that high sound levels may harm one's health, that noise levels that we estimate as innocuous may be dangerous, that there exist an alternative path we can walk to go to work and minimize our exposure to air pollution, etc., citizens will be able to understand the environment around them and act consequently to go toward a more sustainable world.

In order to allow the emergence of the awareness the learning process must take place at a social level, involving individuals both alone and collectively. Participatory sensing, also referred to as urban sensing, involves enabling individuals, groups and communities to gather, document, view, share, and in some cases analyse local observations and data about their surrounding environment. Not all participatory sensing relies on mobile technologies. For example, Haklay et al. [1] comment on the use of low cost noise monitors in a citizen science project in which two communities collected noise data: one in relation to noise nuisance being generated by a local scrap yard and the other, in an objection to an airport expansion plan. However, the use of smartphones as sensory devices, either passively or actively, increases the ability to scale such activities. Cuff et al. [2] highlight a range of applications in which citizens can be engaged in mobile sensing, predicting a growth in the field and in the numbers of ways in which it will be applied.

The power of the crowd has been recognised as an effective way of generating observations, which might otherwise be difficult to obtain, due to spatial and temporal limitations. This is particularly relevant in fields where traditional sensing relies either on a distributed network of expensive stationary monitoring devices across a target area of interest, or where sensors require physical placement for a specific deployment, or in cases where numerical simulations are needed. Cost and data coverage are key factors. The spatial distribution of static monitoring devices and the associated costs of hiring trained specialists to take measurements and process data reduce the amount of real-world measurements that can be taken. That is why, in the EveryAware project, the two main environemental issues faced, i.e. noise and air pollution, have been approached exploiting a crowd-sourcing strategy. The help of volunteers reduces the hiring costs in a significant way, making unnecessary to hire specialist of air pollution monitoring.

Noise pollution is a problem in cities across the world and is one that is likely to affect an increasing number of people with the majority of the global population now living in urban areas, like the World Health Organization reports [3]. In Europe, this has been recognised and abatement measures have been introduced in many countries. However, noise pollution, in particular, is an environmental problem that relies heavily on ‘top down’ approaches, both in terms of communicating the issue, through instruments such as strategic noise maps, but also in the methods used to gather data. For example, strategic noise mapping became a requirement of all Member States under the EU’s European Noise Directive (ENDS). The maps are used to estimate population exposure to noise in certain areas, to communicate to the public and as a basis for action plans, as stated in Directive 2002/49/EC of the European Parliament [4].

Exposure to noise is not merely a case of annoyance. Researchers have provided a growing body of evidence that suggests that long-term exposure to noise constitutes a health risk hazard and can modify social behaviour, cause annoyance (Passchier et al. [5]), increase the risk of cardiovascular diseases (Babisch et al. [6]) and adversely affect levels of attentiveness and the ability to read in children (Haiunes et al.[7]). The World Health Organisation (WHO) estimated that at least one million healthy life years are lost every year from traffic-related noise in the western part of Europe (Fritschi et al. [8]).

Air pollution is another issue which has an important effect on our health, with an increasing number of studies showing higher risk of respiratory and cardiovascular diseases for people exposed to higher pollution levels, e.g. in Lave et al. [9]. In this context, keeping air pollution at bay has been a major priority for policy makers in the past decades. Lots of efforts have been done in monitoring and controlling air pollution. Large scale monitoring networks routinely monitor pollutants. They allow to follow up temporal trends in air pollution. Significant efforts have also been made to make information accessible to the broad public. However, several papers indicate that official monitoring networks do not have sufficient spatial coverage to provide detailed information on personal exposure of people, as for some pollutants, this may vary substantially among micro-environments, like reported in Dons et al. and Kaur et al.[10, 11], i.e. in urban, traffic-prone areas where spatial variability is very high (Peters et al. and Setton et al. [12, 13]). Several pollution sources have been addressed with success. However, persistent problems remain in urban areas, where traffic and domestic heating are important sources, like stated in the European Environment Agency report[14]. Next to the technical solutions (e.g. electrical mobility), people's personal perceptions, behaviour and choices play a major role in addressing these issues and to facilitate change in a bottom-up manner.

In the EveryAware project we addressed to these two main environmental challenge with an aim far more complex then just measuring pollution exploiting the power of the crowd. The goal of our work was the improvement of the involved crowd awareness about those environmental issues and the analysis and the modeling of the dynamics of this improvement. In this chapter we present results from participatory sensing performed using the WideNoise and AirProbe applications and the EveryAware sensor box. We exploit objective and subjective data to provide an analysis of user behaviour/opinions and environmental awareness. In particular, we report on data collected during two large scale test cases: the Heathrow noise pollution test case, organised in London (UK) and the AirProbe International Challenge (APIC) [15], organised simultaneously in four cities: Antwerp (Belgium), Kassel (Germany), London (UK) and Turin (Italy).

2.2 The noise test case

The implementation of the noise test case has already been described in chapter by Theunis et al., in chapter by Atzmueller et al. and in chapter by Nold et al. in this volume. By means of the subjective data collected during measurements an analysis of users awareness will be presented in the following. Subjective data, gathered thanks to the WideNoise app, consists essentially of guesses of the noise level, tag annotation and perception annotation (love-hate, calm-hectic, alone-social, natural-man made) performed contextually with the measure. Widenoise application allows its users to guess the noise level before the actual measurement with the help of a slider ranging from 0 dB to 120 dB. Also, the choice of the noise level measured can be considered a subjective data. The interest is in assessing whether usage of the application leads to any change in behaviour, and whether this change indicates an increase in awareness of environmental noise and its effects. For this study, only data collected by users not belonging to the EveryAware consortium is considered (38267 measurements).

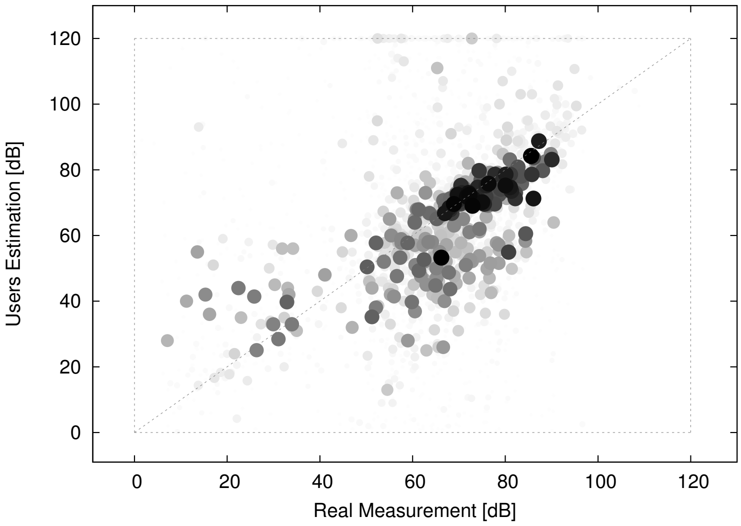

Fig. 2.1 Estimated versus measured noise. Each point corresponds to one measurement, while both the colour scale light to dark grey and the point size represent the user expertise (the first measurement of a given user is depicted with the smallest and lightest point, and are almost invisible; points get darker and larger as users go on with their measures). The graph shows how, with the experience, users become more precise. In fact, larger and darker points, which represent more experienced users guesses, are closer and closer to the measured value.

A first analysis of awareness/learning involves studying the decibel values estimated by users, in comparison with the measured values. Figure 2.1 displays the estimated vs real noise level, with light-coloured small points corresponding to early measurements by a single user, while dark large points corresponding to later measurements. Hence, the size and darkness of points displays user expertise. The figure shows larger darker points closer to the diagonal compared to lighter ones, which means that the estimation is closer to the measured value for later measurements. This indicates that during repeated usage of the application the ability of users to guess the noise level around them increases, hence the user learns in time.

To emphasise this point, Figure 2.2 shows the difference between the estimated and the real noise level as the users repeatedly perform measurements. Averages and standard deviations are also displayed. This shows that as the expertise increases (number of measurements by the same user - horizontal axis), the errors become closer to zero and deviations from the mean decrease.

Considering this, it would be also interesting to see what range of noise is typically measured, and whether this changes in time. Figure 2.3 displays the distri-

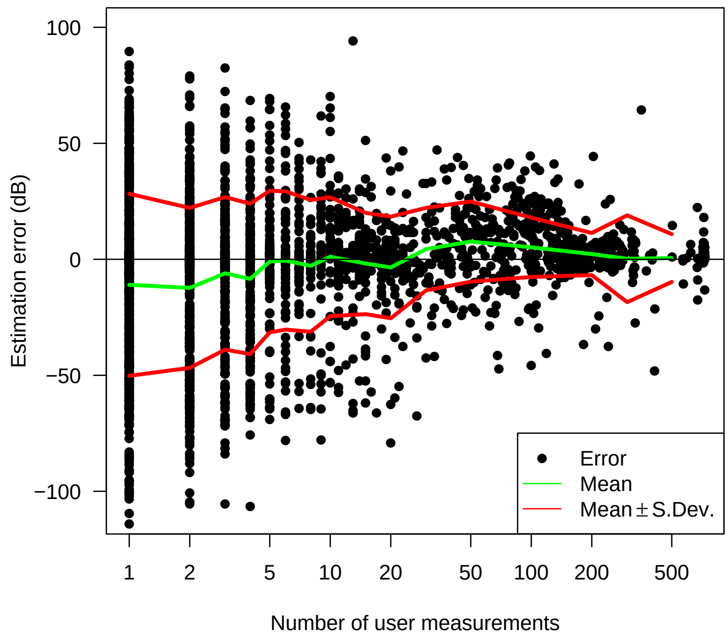

Fig. 2.2 Estimation error. Difference between estimated and real dB value vs the number of measurements a user has performed.

bution of noise levels recorded by all users during their first five measurements, compared to those submitted after having already made 50 measurements (43 users have submitted at least 50 measurements). This shows that the noise levels of experienced users are higher than those of novices, indicating that as users become more involved in measurements they tend to concentrate more on areas with high noise levels, or viceversa users living in noisy areas become more involved in measurements. This could be on one side due to the users learning how to estimate the higher levels of noise, but also due to an increased interest in documenting higher levels of noise in their area.

A different indicator of user involvement and hence awareness is the amount of tags submitted by users. An increase in repeated application usage would indicate increased involvement in data collection and hence increased awareness. Figure 2.4 displays the average number of tags per measurement, considering all measurements submitted to the platform, for increasing level of expertise (measurement number). At the same time, the number of users who have passed a certain expertise level is displayed. This shows that as the users perform more measurements, although the

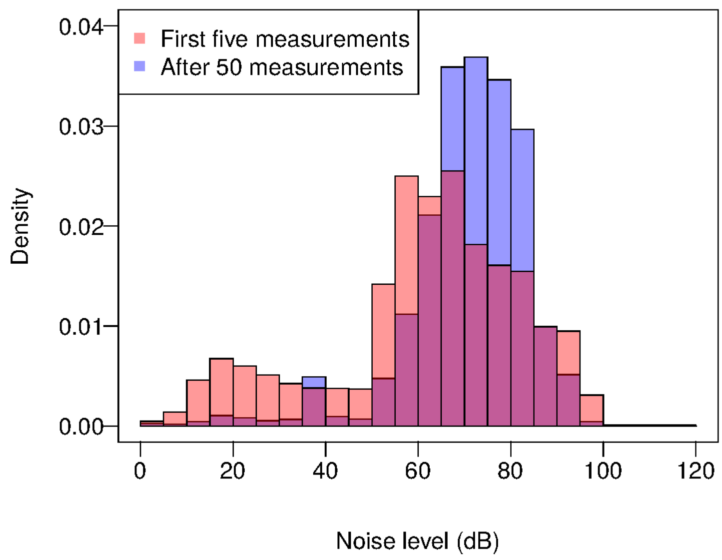

Fig. 2.3 Distribution of measured noise levels. The plot shows the histogram of noise levels for the first measurements performed by users, compared to those performed after some experience is gained (after the 50th measurement).

number of users here decreases, the average number of tags per measurement tends to increase. This demonstrates an increase in user involvement and dedication to the task, hence in the level of awareness.

A further analysis aims to compare the subjective perceptions (Love-Hate, Calm-Hectic, Nature-Man Made, Alone-Social) of the users with the measured noise levels. Out of all measurements performed, 12129 contain intentional perception data. We considered perception data as intended if at least one of the sliders was moved from the default position (0.5). Figure 2.5, shows how these perceptions depend on the measured noise levels. As expected, the perception values increase with noise. This means that, in general, users 'Love' quiet places, finding them a 'Calm' environment, while they 'Hate' loud ones finding them 'Hectic'. At the same time, high levels of noise are in general associated with Man-Made and Social environments.

To analyse the change in opinion as the user is exposed to the information from the application, i.e. the real noise level, Figure 2.5 includes two curves. One shows average perception levels for the first 5 measurements of every user, as a function of noise, while the other shows perceptions for measurements performed after some expertise has been gathered, i.e. more than 50 measurements. The two curves seem

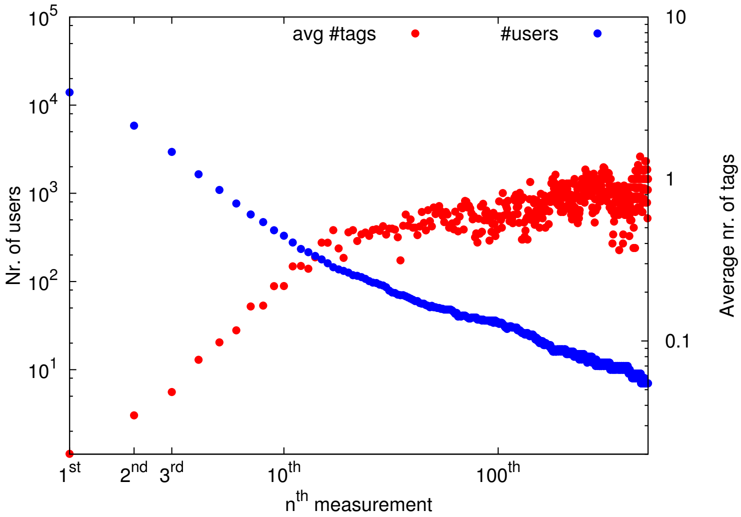

Fig. 2.4 Tagged measurements for different expertise levels. The total number of users submitting at least n measurements is displayed in blue (left axis legend), while the red points represent the average number of tags used in the n-th users' measurement (right axis legend). Both axis are logarithmic.

to show a different behaviour for novice and expert users, for all perception types except for the Alone-Social evaluation. Specifically, noisy environments appear to be perceived as less pleasant and more artificial as the users become more experienced, while quiet environments as more natural and lovable. A sharper switch between the two possibilities is observed around 55-60 dB, for all three types of perceptions, indicating this as a threshold where noise becomes bothersome. This might mean that indeed, exposure to information from the noise application does influence the way in which users perceive the environment. Experienced users have a more stringent evaluation of their environment, and stronger opinions about how much they love or hate the noise levels around. A categorisation of the noise levels appears to emerge, with plateaus visible for high and low levels of noise, when considering data from experienced users. Although it cannot be excluded that experienced users might push the sliders to the extreme right or left edges so to minimize the cognitive effort inherent in judging the quality of noise, the voluntary act of modifying the slider position, by setting it away from the neutral central position, indicates the willingness in conveying a useful information. In that case, we would interpret the pushing of the sliders to the extremes as a conscious act of categorization of experienced users who got more confident with the App. As for the nature-man made indicator, we note that the typical user of our App lives in an urban environment (all the main cast study happened in urban environement), so that there are fewer samples collected in a natural environment and the error bars associated with the

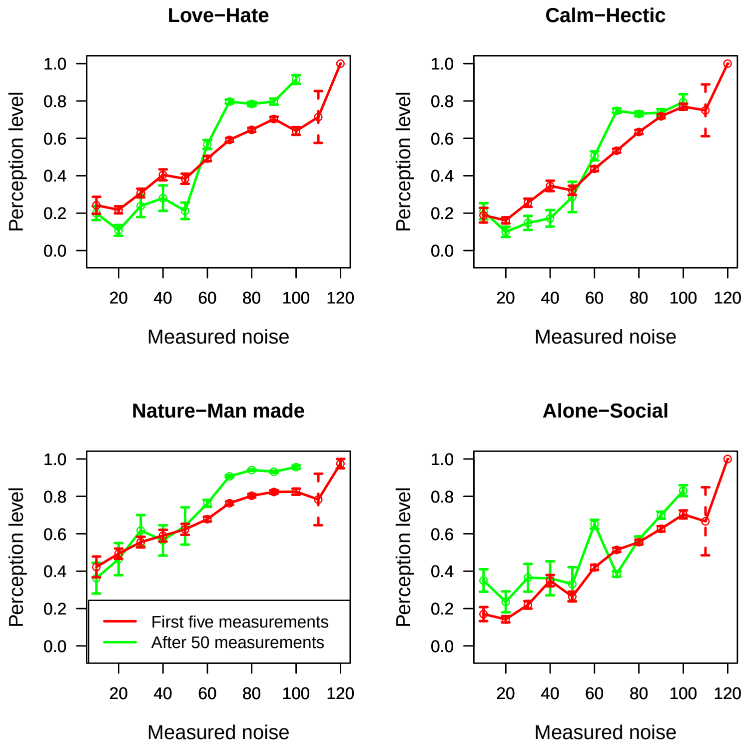

Fig. 2.5 Perception evaluation versus the measured noise level. The red lines display the average evaluation over the first five measurements of all users; the green lines correspond to the average evaluation over the set of all measures taken by users starting from the 50th one.

measures are consequently larger, possibly hiding the categorization effect seen in the other indicators at low dB values. The social aspect, however, does not change with repeated usage of the application, since knowing the noise levels does not affect the user's perception of how many individuals there are around. This explains why there is no definite difference between the two curves in Figure 2.5, lower right pane.

2.3 The air quality test case

During this test case, volunteer participants were asked to get involved in two activity types. One consisted in using a sensing device (Sensor Box, which has been introduced in chapter by Theunis et al. in this volume), to measure air pollution (black carbon (BC) concentrations) in their daily life, generating what we call objective data. The second activity was playing a web game (AirProbe), where volunteers were asked to estimate the pollution level in their cities, by placing flags (so called AirPins) on a map and tagging them with estimated black carbon (BC) concentrations on a scale from 0 to 10 $\mu g/m^3$, resulting in subjective data on air pollution (perception). Volunteers involved in the measuring activities were also encouraged to play the game and bring other players as well.

The two data types allow for an analysis of user behaviour and perception throughout the challenge. To enable this, the test case was composed of three phases. In phase I, only the online game was available, so we could obtain an initial map of the perceived air pollution. In phase II the measurements started in a predefined area in each of the cities (corresponding also to the game area), with the web game running in parallel. Phase III introduced a change in the game, so that players could purchase information about the real pollution in their cities. At the same time, measurements were continued, this time without a restriction of the area to be mapped.

Volunteer involvement and activity levels are among the most important elements in participatory monitoring campaigns, since these can decide the faith of entire project. Minimal activity is required for acquiring data, both objective, for analysis of the environment itself, and subjective, for analysis of social behaviour. The test case presented here has successfully involved 39 teams of volunteers in 4 european locations, gathering 6,615,409 valid geolocalised data points during the challenge (the measuring device collects one data point per second). An additional 3,326,956 data points were uploaded to our servers in the same period, but missing complete GPS information, so were not included in the analysis. Some of these measurements contained labels (tags), with 742 geo-localised tags coming mostly from one location of the challenge (London).

Additional information on perception of pollution has been extracted from the online game. The platform had 325 users in total, over six weeks, 97 of which played the game at least ten times. Their activity resulted in 70,758 evaluations of pollution (AirPins) at the end of the test case. However, some other AirPins had been added or values had been modified during the challenge, so that the entire data used was much larger.

For insight into measurement coverage patterns and how these evolved during the test case, Figure 2.6 displays coverage in space obtained every week, together with the overlaps between the different weeks. Space coverage is computed by dividing the area of each of the four participating cities into 10 by 10 meter squares (tiles). One square was considered covered if at least one measurement was performed within its area. Overlaps are obtained through the intersection of covered tiles in different weeks. Both overall values (use entire dataset to mark tiles that are covered or not), and team averages (compute coverage and overlap for each team then average over all) are displayed. The former provide insight into the quality of the dataset obtained, while the latter indicate measuring strategies.

Overall coverage shows that every week all volunteers mapped more than 5 km^2, with higher values in the first two weeks. This is probably due to the fact that in these two weeks they were instructed to cover as much as possible from a specific area, while in the second fortnight they were asked to use the sensor box how they wished. Pairwise comparison of the different weeks shows over 30% of the area is covered in at least two weeks. The overlap between the first two weeks reaches over 50%, while following weeks have less overlap. This indicates that one can obtain good coverage both in time and space by indicating a restricted area for mapping. Also, this appears to indicate that during the last two weeks of the challenge volunteers explored more, since the overlap between weeks is lower.

To test this hypothesis, we also include averages per team for coverage and overlap in Figure 2.6. Coverage is very high during the second week of the test case and comparable for the rest. This may be because the main prize of the challenge was given for second phase activity, i.e. at the end of the first fortnight of measurements. So, volunteers made an extraordinary effort the week before the prize, after a first week of exploration. Overlap on the other hand gives opposite indications compared to overall values. The highest overlap, of about 20%, is seen during the last two weeks of measurements. This means that volunteers make more measurements on the same path than in the first two weeks, so they explore less. This indicates that while in the first weeks they explore wider areas because of the incentives, when these are removed they reduce the area of interest, probably to most familiar and frequented locations. The overall values (top-right panel of Figure 2.6) seemed to indicate more space exploration during the last phase, but this was an artefact of the fact that the area was restricted in the first two weeks, so overlap between volunteers was much higher, increasing the overall overlap as well.

The measured BC levels can also provide useful insight into the aims and strategies of the volunteers during the challenge. The two measuring phases (phase 2 and 3 of the test case) gave different tasks for the volunteers. In phase 2, they had to concentrate on covering as much as possible a specific area, while in phase 3 they could explore any area they wanted. It would be interesting to understand if the measured BC levels changed between the two phases. Of course, pollution levels themselves may change from one day or period to another. In order to measure the change in BC levels due to change in behaviour and not due to actual changes in the pollution levels, we need reference pollution data for the days of the challenge. For all four locations, average daily PM10 (particulate matter smaller than 10 micrometers) values were obtained from public repositories and used as a baseline for normalisation. BC levels were not available for the same locations, however PM10 correlates very well with BC levels, so can be used also as a baseline (in general, PM10 concentrations are more or less 10 times larger than BC levels, e.g. like reported in Vanderstraeten et al. [16]). These daily averages were used to scale all measures performed by our volunteers. In the following only these normalised BC levels will be used to build the discussion on real measurements.

Figure 2.7 shows histograms of normalised BC levels measured in the two phases, and we can observe larger BC values in phase 3. One could argue, in this situation, that probably most of the measurements in phase 2 were within the monitoring area, which we selected in the city centre, where limited traffic zones exist, so that could explain the difference in BC levels between the two phases. This is why we show data from within and outside the monitoring areas separately. The increase

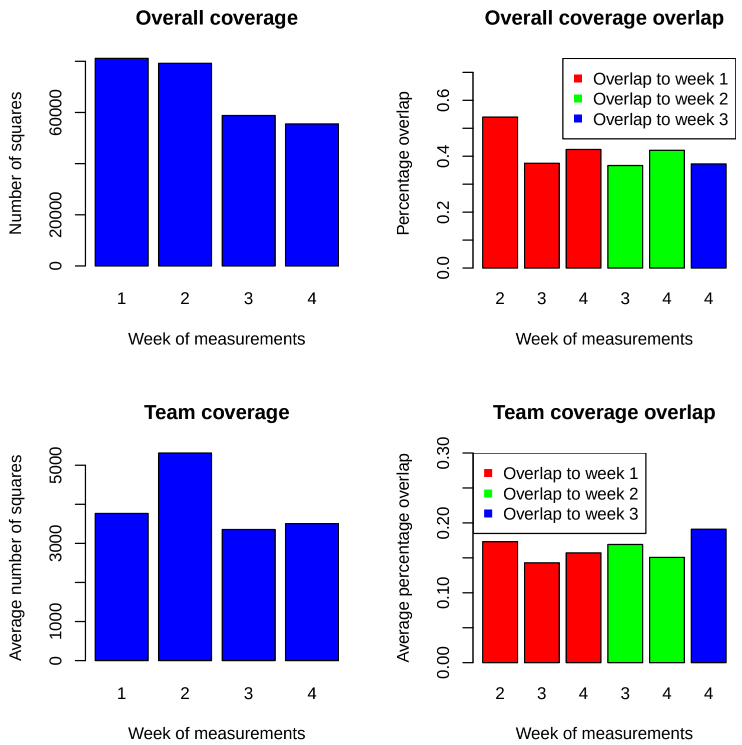

Fig. 2.6 Coverage per week and overlap between weeks.

in BC levels in visible for both cases, so we believe it is due to the interests of the volunteers, and does not depend on the area to be monitored. When they can choose freely where to make measurements, volunteers appear to be driven to trafficked more polluted areas, since it is those locations what they want to identify first.

To look into this even further, Figure 2.8 shows the distribution of normalised BC for the different locations, compared in the different phases. Again, data inside and outside the monitoring areas is shown separately, and the box width highlights the significance of the normalized value based on the size of the averaged set. In Kassel, volunteers were grouped into two groups in phase 3: the first group (g1 -three users) had as a task to avoid highly polluted areas, while group g2 had no task other than using the sensor box where they wished. This, in order to test whether any learning appears during measurements.

For Antwerp, volunteers collected higher BC levels in phase 3, both outside and inside the monitoring area. In London, although means are not larger, the maximum levels achieved are larger in phase 3. However, for these two cities data in phase 3 is rather limited compared to the other locations and to phase 2 (as shown by the width

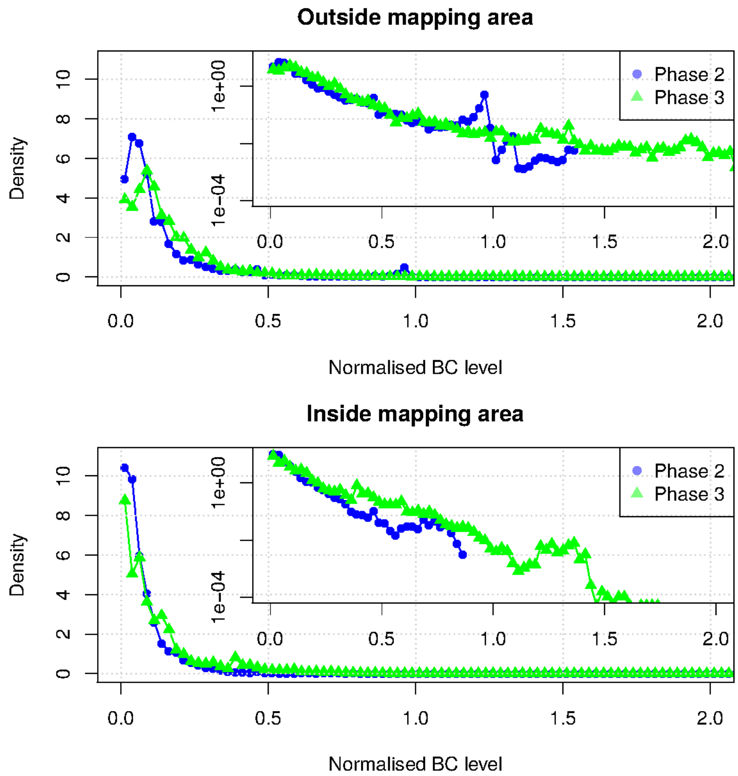

Fig. 2.7 Overall pollution levels compared between the two phases. BC levels normalised by scaling with average daily PM10 concentrations are shown for the two measuring phases of the challenge. The data measured within the monitoring area of phase 2 is considered separately from that measured outside, to control for the different setting. Inset graphs logarithmic version of the plots.

of the boxplots in figure 2.8). For Turin, an increase in the measured pollution levels is clear outside the monitoring area, but not visible inside. So, for all three locations, there is a good indication that volunteers concentrated more on high pollution levels in the 3rd phase of the challenge: when they were allowed to explore, the aim was to identify highly polluted locations.

For Kassel, the group supposed to minimise their exposure displays lower BC levels compared to the other group only inside the monitoring area, while outside this they measure higher pollution levels. Maximum values appear, however, to be lower than the previous phase. This indicates that volunteers have successfully learned how to avoid high pollution levels within the monitoring area, after two weeks of exploration. However, they are not fully able to extrapolate this knowledge to unseen locations, although they do manage to avoid very high pollution spots.

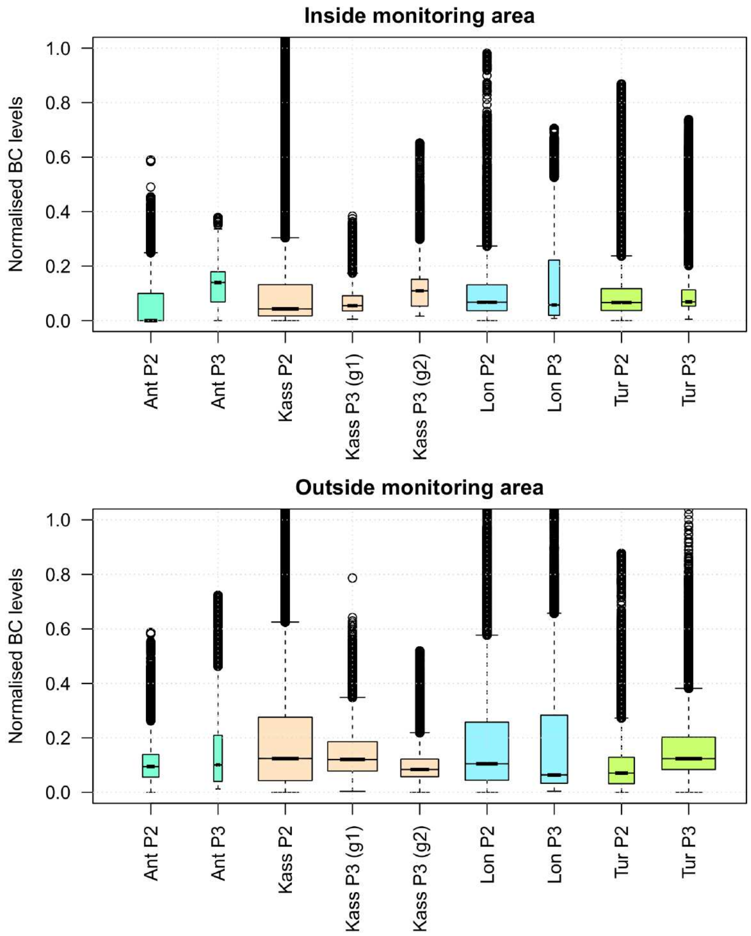

Fig. 2.8 Pollution levels per location compared in the two phases. The distribution of BC levels, normalised by scaling with average daily PM10 concentrations, are shown for the two measuring phases of the challenge, separate for each location. The data measured within the monitoring area of phase 2 is considered separately from that measured outside, to control for the different setting. Width is an indication of the size of the dataset.

One question is why the exploratory behaviour, keen on higher pollution levels, seen in phase 3, when volunteers are free to use the sensor box where they want, does not also appear in phase 2. A possibility is that the exploration does happen at the beginning of the phase. However, given that the area is restricted, this stops after some time and afterwards the only aim remaining is covering the area. To check this, we have looked at average normalised BC every hour of measurements, for each user, and then averaged this over all users. Figure 2.9 shows the values obtained in the two phases. It is important to note that here the time axis represents user experience: the first point represents an average over the first hour of measurements for all users, the second the average over the second hour of measurements, even these may have happened at totally different times for each user. For instance, if a user decided to start their activity on the second day of the test case, then their first hour will be one day later than the other volunteers. For this reason, as the number of hours increases, the number of users that have reached that level of experience decreases, and this is also shown in the figure.

Indeed, measurements made in the first hours of sensor box usage, in phase 2, yield larger BC levels, indicating that at first volunteers looked for highly polluted spots. As they become more experienced with the box, and they identify more such locations, the BC values they measure decrease slowly (although fluctuations remain), indicating a loss of the exploring interests. This could also indicate volunteers are learning how to avoid highly polluted spots. The same patters is preserved if volunteers with a low total number of measurements are excluded from the beginning from this analysis. Another possible interpretation implies the presence of some users who are concerned about air pollution problem. So they tend to avoid pollution and, since they are really engaged (because of their concerns) they measure air quality longer. So at the end the air pollution line goes lower.

For the third phase, however, no decrease is visible in the measured BC levels, until the number of users becomes very low (2), where fluctuations may be due to local variability so are not relevant. Hence, indications are that during this phase users continued their exploration for the entire two weeks, since there was no limitation on the area to be covered.

The analysis of the structure and location of the collected objective data gives some insight into what volunteers are interested to see when measuring air pollution and whether any learning appears. Subjective data, on the other hand, can provide a stronger indication of changes in perception. For this, we look at data collected through the web game, which consists of perceived levels of pollution geolocalized in the mapping area. These were obtained by asking players to place on the map AirPins, geolocalized guesses of the air pollution levels. Figure 2.10 shows the distribution of the perceived pollution at the end of each phase of the challenge.

Data from the first phase represent the original perception of air pollution by the volunteers: during this phase, players had no access to sensing devices nor any data. The distribution of pollution levels appears to be bimodal, which is an indication of a categorization effect. Volunteers divide the locations into those with very low pollution and those with higher pollution. The higher pollution levels peak around the middle of the pollution range, with larger and smaller values also present. This indicates that players took the middle of the range as a medium pollution level and moved around this to tag the different locations in the city.

In the second phase, however, some volunteers were given the sensor boxes to start performing measurements. The web game players consisted of these volunteers plus a set of other players recruited by them, so from their friend circle. No data, except for the direct feedback from the boxes, was shown to the volunteers. Even so, a change is visible in the distribution of perceived pollution levels reported in the web game. Volunteers see that in general BC concentrations are lower than what they

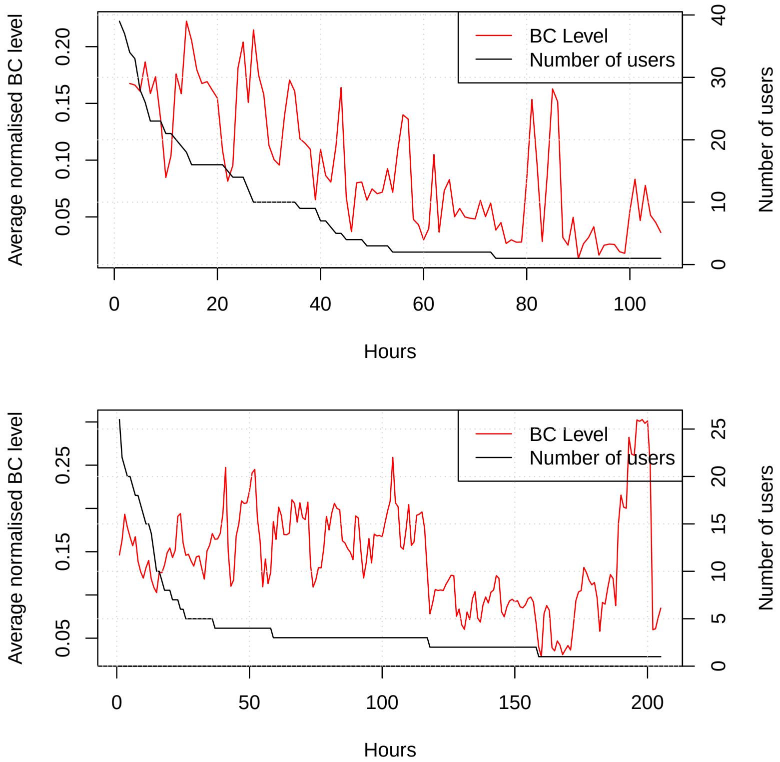

Fig. 2.9 Pollution levels and user experience. Average normalised BC levels are shown for all users, by considering one hour of measurements at a time. Hence, the horizontal axis can be viewed as user experience, i.e. how many measurements they have performed, with the red line showing how the hourly BC level changes as the user makes more measurements. The black line shows the number of users that reached a certain experience level (for instance, in the top panel, only two users performed 60 hours of measurements, so only their data is displayed). The top panel corresponds to phase 2, while the bottom panel to phase 3.

believed, and respond by changing the values of the AirPins. Since the change is quite significant, we also believe that those volunteers with the sensor boxes spread the information about what they were measuring, so that all players changed their perception. This decrease in the pollution levels reported in the subjective data of phase two is a very strong indication of learning during this phase.

In phase three, perceived pollution levels decrease even further. However, here the mechanism is different. Players are now allowed to purchase information about average pollution in different map tiles (called AirSquare), so they can now adjust their guessed pollution levels based on that. So, in this case the change is triggered from within the game, while in phase two the change appeared naturally from the user experience outside the game.

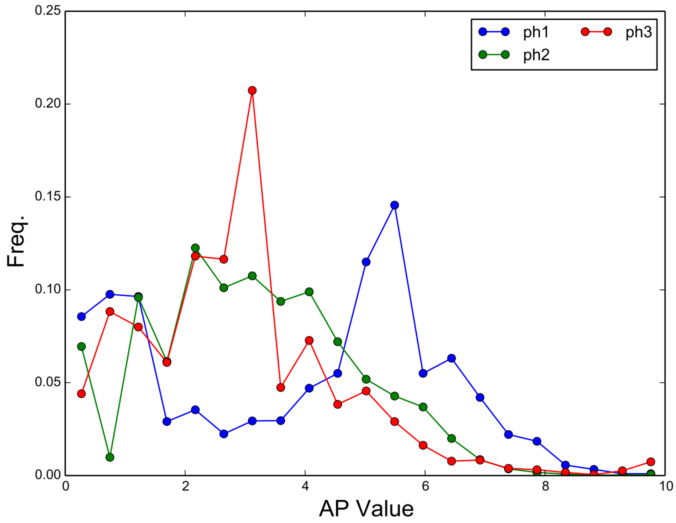

Fig. 2.10 Web game subjective data. The plot shows the distribution of perceived pollution levels (AirPin values) collected at the end of each phase of the test case.

2.4 Emergence of awareness in the AirProbe Web-game

By playing the AirProbe web-game users are exposed, in phase III, to the air quality measures collected by the Air Ambassadors (volunteers equipped with sensor boxes). Therefore they are somehow learning the air quality status of their environment. However, it is not well justified to assume that what is learned by players within the game is equivalent to awareness. Awareness is a slow process with long characteristic time scales so that it is not feasible to measure it in a short lived experiment as this one. Nevertheless, we can try to understand whether, in the game context, the behaviour of players differs from the trivial task of setting AirPins (AP) values just by copying the value shown by the purchased AirSquares (AS). If any systematic difference is detectable we could ascribe it to a sort of an opinion shift toward a virtual awareness. To this aim we shall report here the evolution of the difference between the AP value and the value of the AS it belongs to. This difference will be referred to as AP difference (APD) in the following and is displayed as heat maps in Fig. 2.11, Fig. 2.12 and in Fig. 2.13 for the city of Kassel, London and Turin, respectively. Antwerp dataset was discarded because of its negligible size.
Once more, we observe the effect of the overrating due to the wrong scale usage in the first two phases. Interestingly, the maps related to phase III indicate that players tend to overestimate the values in those places that were previously annotated as very polluted. We will analyze this kind of effect in detail in the following.

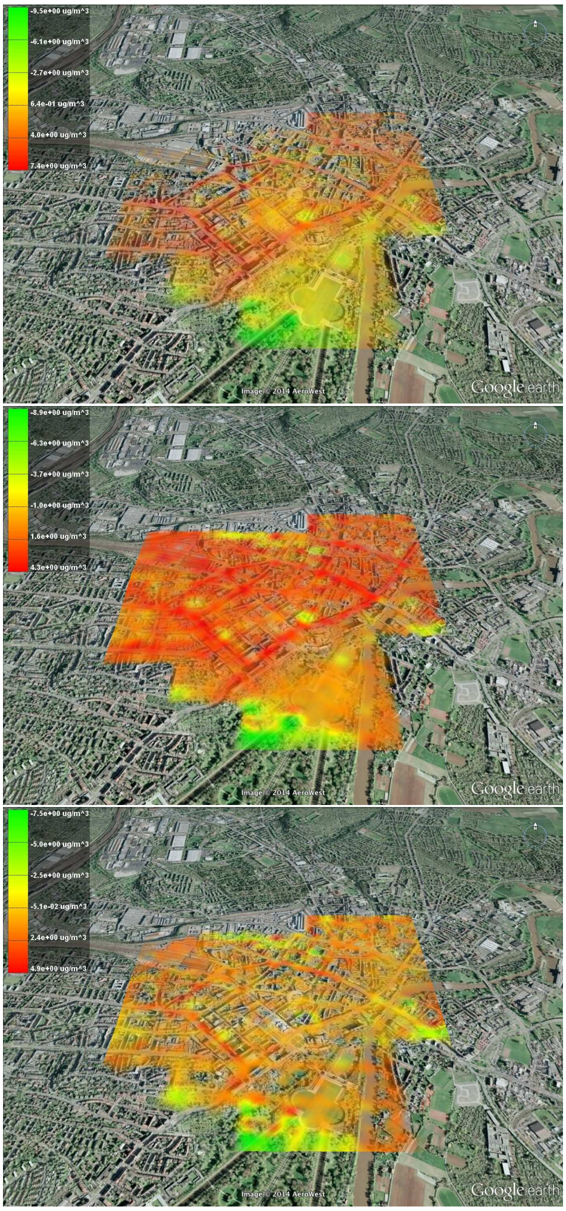

Fig. 2.11 Heat map of the APD values (difference between the web-game annotated AirPin value and the AirSquare value inferred from on field sensor box measurements) for the city of Kassel. Top figure refer to phase I, the middle one to phase II, the figure at bottom to phase III. Values in the legends represent $\mu g/m^3$ concentration of Black Carbon. The opacity is related to the number of AirPins used in the corresponding point.

Fig. 2.12 Heat map of the APD values (difference between the web-game annotated AirPin value and the AirSquare value inferred from on field sensor box measurements) for the city of London. Top figure refer to phase I, the middle one to phase II, the figure at bottom to phase III. Values in the legends represent $\mu g/m^3$ concentration of Black Carbon. The opacity is related to the number of AirPins used in the corresponding point.

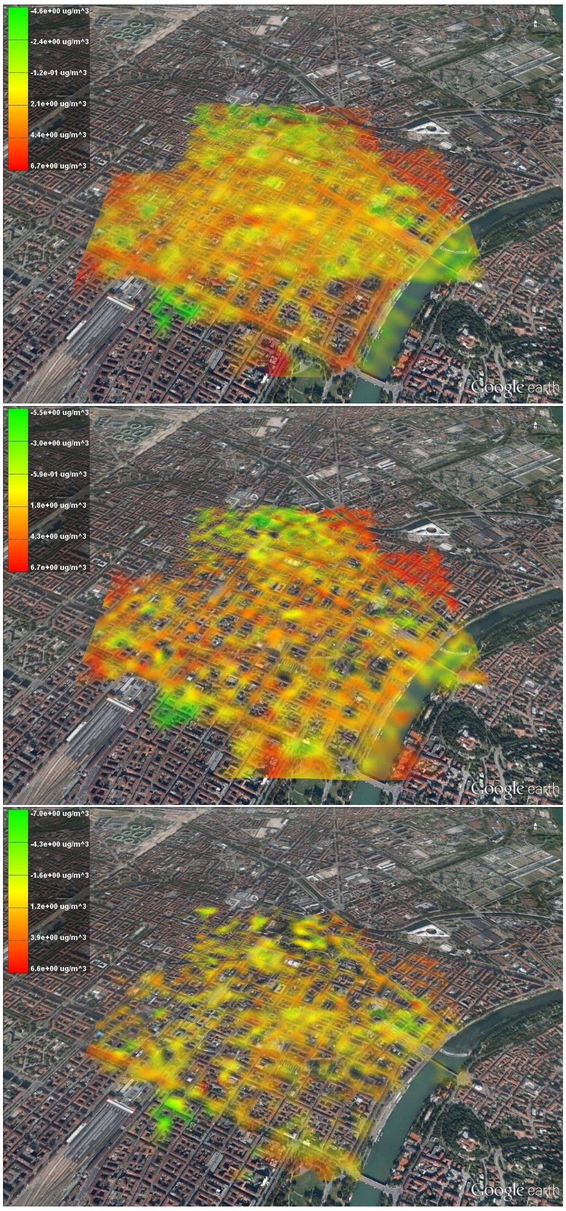

Fig. 2.13 Heat map of the APD values (difference between the web-game annotated AirPin value and the AirSquare value inferred from on field sensor box measurements) for the city of Turin. Top figure refer to phase I, the middle one to phase II, the figure at bottom to phase III. Values in the legends represent $\mu g/m^3$ concentration of Black Carbon. The opacity is related to the number of AirPins used in the corresponding point.

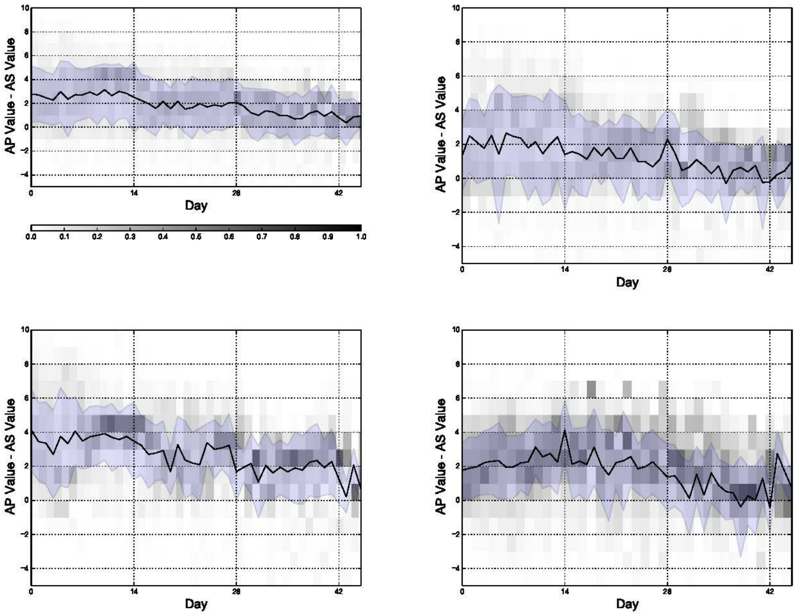

Fig. 2.14 Distribution of the APD values in time. Each column displays the histogram of APD in the given day with each bin painted in a gray scale level related to the relative importance of the bin (white means no APDs fall in the bin, black means all APDs fall in the bin). Bin size is $0.5\,\mu g/m^3$. The curve shows the average daily value while the blue area the corresponding standard deviation. The plot at the top left is calculated for the overall set of data, while, going on clockwise, the other plots refer to Kassel, Turin and London respectively.

In order to understand what is going on here time is a key factor. Thus we measured the evolution of the opinion with the histogram of APD daily values reported in Fig. 2.14, where we also added a line showing the daily average and a bluish region depicting the corresponding standard deviation. Overall, players are overestimating the pollution of their environment, though it is not clear whether this is a result of being rather pessimistic or of not having correctly grasped the scale used to report the air quality parameter chosen. After each change of phase, i.e. at day 14 and 28, a major shift of APD can be spotted (except in the case of London at day 14). In each shift, the APD decreases, showing that people begin to understand better the black carbon scale used in the game and are improving their evaluations. At day 14, i.e. at the switch between phase I and II, Air Ambassadors started their measuring activity and sharing information with Air Guardians (players of the web-game) of their teams. Moreover, at day 14 some rules of the game changed by stimulating players to be more precise in their estimations. This kind of transition seems to be quite fast, since the shift takes only few days both in Kassel and in Turin (in London there is a slightly different situation). The substantial steadiness of the APDs along the duration of each phase allow us to consider phase-aggregated data in order to answer to the original questions: how does the shift take place? Are volunteers learning something from the game or are they just blindly copying the AS values?

Let us now look at the APD histograms aggregated according to each phase. Since the time scale for opinion shift seems to be very short and the opinion distribution seemed to be more or less constant, data aggregation by phase sounds reasonable. We are interested in how the exposure to information affects opinions, so we will consider only those APDs for those AirPins whose relative AirSquare was effectively purchased by the user. The assumption about the opinion stability during each phase is particularly important in phase III. This implies that in the last phase players bought a great number of AirSquares in the first days and in those days their opinions changed. So we can consider all AirPins of phase III as projections of the opinion shifted as a consequence of the exposure to the AirSquare information. How this reflects on the APD distribution is reported in Fig. 2.15.

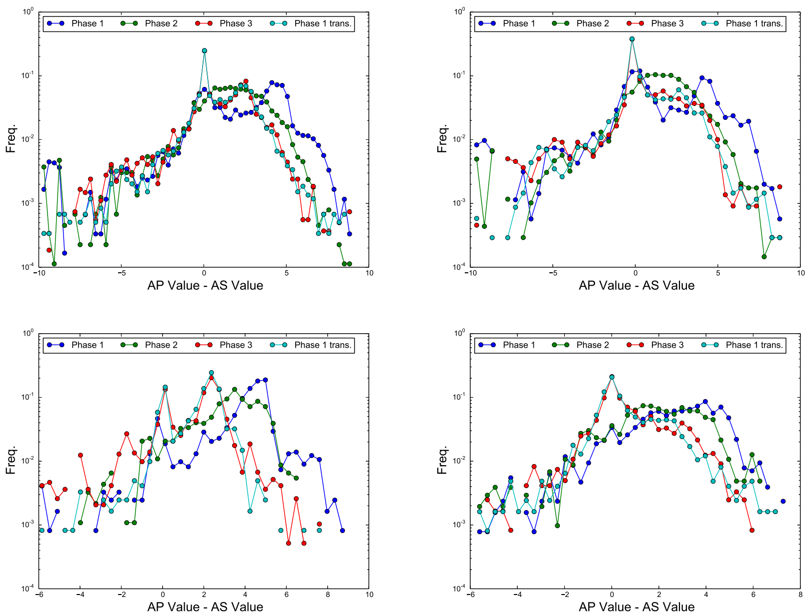

Fig. 2.15 Clockwise, from the top left: the APD histogram for the overall, for Kassel, for Turin and for London in each phase of the challenge and with an estimation of phase III data obtained from phase I data through the transformation defined in Eq. 2.2.

If we look at phase three histograms two main features attract our attention: a narrow peak in 0 and a deeply asymmetric structure. The first feature was somehow expected since players are trusting the AS values shown in the AS, and they are annotating accordingly. Fortunately, the peak at zero is not delta like, what is expected for users copying the AS value. Rather players still have their opinion on the environment and keep it despite the on field measurements. This may happen because they are really trying to follow the basic ideas of the game but also because copying it is not the best strategy, since they know that the AS value is aggregated, i.e. it is the average of all sensor-box measures taken in the corresponding AS, while the real measurements used for revenue calculation were punctual values which could be substantially different. So the shape of the distribution around zero seems to be caused by users learning the most likely air quality value and trying to estimate fluctuations. But graphs in Fig. 2.15 show something more. There is a clear asymmetry for phase III distributions, since the great part of APD values fall in the positive range. This could be a consequence of the fact that AS values were around 3 $\mu g/m^3$ so there was a 30% probability to underestimate that value and 70% to over estimate, but if we look at the phase I distributions, this asymmetry effect seems better explained by a sort of memory effect or inertia of players in changing their opinions. This hypothesis seems realistic if we look at the London graph. The main peak around $4\mu g/m^3$ is still present in phase III, although it is shifted. In order to measure this effects we defined a transformation that takes into the account both features just discussed: the accumulation around 0 and the shift. Let us consider a given set of opinions o_i about a certain number of topics provided by a certain number of subjects. At a given time those subjects are exposed to values h_i, which are perceived as hints of the true values. We are interested in what happens to the difference between opinions and hints before and after the exposition, to understand how this information will affect the opinion structure. To this aim, we define the set of differences d_i between the opinions and the relative hints and analyse the distribution of those difference before and after the exposition. Obviously, the variation of the differences is only due to the variation of the opinions. As we said, we want to reproduce the phenomenon of the accumulation around the hints (i.e., $d_{aft} \sim 0$) and the shift of the general opinion, that we will try to describe as a sort of rescaling (i.e., $d_{aft} \sim d_{bef}/r$ where r will be the rescaling constant). Which of the two phenomena will take place will be decided randomly: with a given probability p_0 the opinion will reset around 0, otherwise, with probability $1 - p_0$, the opinion will just be rescaled. Finally, around this two attractors we add a certain amount of noise. We decided for a Cauchy distribution $C(X)$ centered in 0 in one case and in d_{bef}/r in the other, i.e.

$$C(x;\mu,\gamma) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-\mu}{\gamma}\right)^2\right)} \tag{2.1}$$

where μ is the average (and the center of this symmetric distribution) and γ represents a scale factor. It is worth to note that the variance of this distribution is not defined, since the second momentum of the distribution does not converge. This choice seems reasonable because tails seem to be power law-like rather then gaussian-like, as the log plots in Fig. 2.15 show. Let us define our transformation and its effect on the difference d_{bef} between the opinion and the hint before the exposure. According to the rules we stated earlier, d_{aft} will be distributed according to this density function: 2.2

$$T(d_{aft}; d_{bef}, p_0, r, \gamma_0, \gamma_r) = \begin{cases} C(d_{aft}; 0, \gamma_0) & \text{with prob. } p_0 \\ C(d_{aft}; d_{bef}/r, \gamma_r) & \text{with prob. } 1 - p_0 \end{cases} \tag{2.2}$$

The transformation we just defined introduces four parameters:

- p_0, which is the probability that the old opinion is reset around $d = 0$; thus, with probability $1 - p_0$, the opinion shows a certain inertia; this resistance to change causes a shift toward the hint instead of a complete reset;
- r, the rescale factor quantifying the shift of resilient opinions;
- γ_0 and γ_r , the γ scale factors for the Cauchy distributions centered respectively in 0 and in d_{bef}/r introduced to add a realistic noise.

We used our data to infer the parameters of our model for Kassel, London, Turin and for the complete set of data. If we apply the transformation to phase I data, we get an estimation of phase III distances between opinions and hints. Then, to evaluate how good is the estimation, we use a two sample Kolmogorov-Smirnov two sided test. This kind of test gives as result the probability p_{val} that the hypothesis that the two samples are drawn from the same distribution cannot be rejected. Usually, a value below 5% means that the hypothesis has to be rejected otherwise the hypothesis is likely to be true. If the p_{val} is around 10% the two samples come from two distribution which are, in any case, very close. Above the 30% the samples can be considered with a good degree of confidence as coming from the same distribution. We explored the space of parameters with 10% steps and repeating the test 100 times to find the combinations with the highest p_{val} for Kassel, London, Turin and for the overall. These optimal combinations are reported in Table 2.1 with the relative results for the Kolmogorov-Smirnov test.

Table 2.1 Parameters combination with the highest p_{val} resulting from the Kolmogorov-Smirnov test. Parameter space has been explored with 10% steps and each configuration has been tested 100 times. The average p_{val} is reported. Some peaks in the tails for London compromised the test, causing as a result unsatisfying values for the parameters. We reduced the range in the most meaningful area, which is $(-1:4])$. We found the best parameters testing only this area, obtaining a remarkable result ($p_{val} = 27\%$). Then we made again the test reintroducing neglected data, obtaining a $p_{val} = 9\%$ which is still a satisfactory result.

dataset	p_0	r	γ_r	γ_0	$< p_{val} >$
Kassel	0.336	1.62	0.381	0.0138	0.192
London	0.147	1.90	0.100	0.030	0.267 (0.087)
Turin	0.583	1.56	0.304	0.300	0.417
Overall	0.204	1.767	0.28	0.015	0.262

From the table seems that the reset of the opinion around the hint happens not so often. In London, for example, it is almost a secondary effect. In the best case, Turin, the reset seems to be there slightly more then in the half of the cases. We also reported in Fig. 2.15 an estimation of the APDs for phase III obtained by applying the transformation 2.2 with the optimal parameters combination to the data of phase I. The similarity between estimation and phase III real data is pretty clear.

It is very likely that Eq. 2.2 is not the real transformation of the opinion due to the subjects exposure to hints. We made strong assumptions and we reduced our data set to focus on the interesting part. Also, we are analyzing and modeling the phenomenon on a very narrow timescale (weeks) without knowing almost anything about the others (for example, if we consider months the dynamics could be potentially extremely different). Despite this considerations, the results we showed is novel, to our knowledge, and seem to point out with sufficient reliability that the main ingredients are there. The model we referred to helped us to measure how our volunteers were influenced by the hints we gave them. We may now affirm with a certain degree of confidence that even when people do not trust completely the AS values, they still get influenced by them. Another way to see this is that, even if people do not reset their opinions, the space itself in which their opinions are arranged is deformed by the exposure to hints. Obviously these considerations are justified if the subjects consider the source of the hints as objective. In other cases, for example if volunteers are told that opinions come from other volunteers, completely different dynamics are expected to come into play.

2.5 Conclusions and Perspectives

Volunteer participation is crucial for the success of bottom-up monitoring campaigns, however most projects concerned with environmental monitoring concentrate still on the development of the technical tools necessary. In the EveryAware project we gave a different user-centric perspective though its large scale test cases for noise and air pollution.

For the noise case, several indicators have been derived from the objective versus subjective data submitted by users, leading to the main findings:

- Guessed levels of noise, compared to the measured ones, indicate that users learn to estimate the noise level after repeated usage of the application.
- Perception rating is shown to change in time, as users perform more measurements. Hence noisy environments are qualified as more hectic and less lovable by experienced users, compared to novices.
- An increase in the fraction of tags submitted by users was observed as these became more experienced. This suggests an increase in involvement and dedication with time. Together with the change in perception, this indicated an increase in awareness after repeated usage of the WideNoise application.

For air quality, objective measurements allowed for analysis of user interests during the challenge, as well as learning. Both coverage and pollution levels measured indicated a tendency to monitor familiar areas when this was not restricted, with a search for highly polluted spots. However, as users become more familiar with an area, the levels of pollution decrease in the data, a first indication of learning how to avoid high pollution levels. Subjective data, on the other hand, allowed for analysis of perceived pollution levels. Volunteers started with a strict categorization in polluted and non-polluted areas, where pollution in affected areas was overestimated. Through usage of the EveryAware sensor box, they however adjusted their image, decreasing overall pollution levels. This shows that involving volunteers in monitoring campaigns can help learning to build a more accurate perception of air quality issues.

By means of a web-based game, the inertia of citizens to change their opinion on the air quality level of the urban environment was estimated. Interestingly, citizens seem to be reluctant, in a statistical sense, to change their opinions that are typically of pessimistic character and stick to their personal feelings rather than to trust data stemming from official measures. We observed, anyway, that these data have a non-trivial effect on citizens opinions, deforming the very space in which opinion are arranged. This information can be of interest for stakeholders and decision makers in order to find new efficient ways to improve awareness..

To the authors knowledge, this is the first study where a throughout parallel investigation of objective and subjective data has been performed, hopefully boosting an increase in awareness toward environmental issues. Beside the value as a proof of concept, the EveryAware project also succeeded in providing meaningful insight about the awareness and opinion shift mechanisms.

Although initial signs of learning and increased awareness have been found already at this level, the usage of the application and evaluation of indicators such as those presented here will be continued in the future. Additionally, an in depth study of several data components is envisioned for future work, such as a semantic analysis of tags, which could give further important insight into both the motivation and opinion of users about their environment.

References

[1] M. Haklay, L. Francis, , C. Whitaker. Noise Mapping Helps Citizens Take Action (2008)

[2] D. Cuff, M. Hansen, J. Kang, Communications of the ACM **51**(3), 24 (2008)

[3] World Health Organization, The WHO Centre for Health Development, Kobe, and United Nations Human Settlements Programme, (2010)

[4] E. Commission, et al., Official Journal of the European Communities L189 of **18**(2002), 12 (2002)

[5] W. Passchier-Vermeer, W.F. Passchier, Environmental health perspectives **108**(Suppl 1), 123 (2000)

[6] W. Babisch, B. Beule, M. Schust, N. Kersten, H. Ising, Epidemiology **16**(1), 33 (2005)

[7] M.M. Haines, S.A. Stansfeld, R.S. Job, B. Berglund, J. Head, International Journal of Epidemiology **30**(4), 839 (2001)

[8] L. Fritschi, L. Brown, R. Kim, D. Schwela, S. Kephalopolous, *Burden of disease from environmental noise: Quantification of healthy life years lost in*

Europe (http://www. who. int/quantifying_ehimpacts/publications/e94888/en/, 2011)
[9] L.B. Lave, E.P. Seskin, *Air pollution and human health*, vol. 6 (Routledge, 2013)
[10] E. Dons, L. Int Panis, M. Van Poppel, J. Theunis, G. Wets, Atmospheric Environment **55**(0), 392 (2012)
[11] S. Kaur, M.J. Nieuwenhuijsen, R.N. Colvile, Atmospheric Environment **41**(23), 4781 (2007)
[12] J. Peters, J. Theunis, M. Van Poppel, P. Berghmans, Aerosol and Air Quality Research **13**, 509 (2013)
[13] E. Setton, J.D. Marshall, M. Brauer, K.R. Lundquist, P. Hystad, P. Keller, D. Cloutier-Fisher, Journal of Exposure Science and Environmental Epidemiology **21**(1), 42 (2011)
[14] European Environment Agency, Air quality in Europe - 2013 report. Tech. rep. (2013)
[15] AirProbe International Challenge. www.everyaware.eu/category/apic/ (2013)
[16] P. Vanderstraeten, M. Forton, O. Brasseur, Z.Y. Offer, JEP **02**(05), 525 (2011). DOI 10.4236/jep.2011.25060. URL `http://dx.doi.org/10.4236/jep.2011.25060`