Challenges and Limitations for the Systematic Collection of Cycling Data from Bike Sensors

Miguel Costa and Rui José(✉)

Centro Algoritmi, University of Minho, Guimarães, Portugal
rui@dsi.uminho.pt

Abstract. Information and Communication Technology is increasingly recognised as a key element for the ability of cycling mobility initiatives to create real, profound, incremental and measurable impact. Even though previous work has extensively explored many applications of smart cycling data, the first challenge is to actually produce consistent cycling data in a systematic way. In this research, we explore the range of sensors which could be more relevant to integrate into urban bicycles to support the systematic collection of data about cycle routes. To gain a deeper insight into the real-world challenges of systematic cycle-based sensing, we conducted an experimental data collection. We equipped a bicycle with a diverse set of low-cost sensors, and we collected data in a pre-defined route, in which it was possible to experience very diverse environmental circumstances regarding road surface or the level of surrounding traffic. The results highlight some of the practical challenges that can be faced by systematic sensing for urban cycling, suggesting that not all sensors might be appropriate for this type of large-scale deployment on bicycles. The main contribution is a set of design implications, which should help to inform the design of novel sensing systems for bicycles.

Keywords: Bicycle sensing · Smart cycling · Mobility data

1 Introduction

Cycling is becoming a key element in smart mobility policies [1]. This new reality is emerging in a context of sustainability agendas, but also as a fundamental path towards more liveable cities, where public space is rescued from cars and given back to citizens. Information and Communication Technology (ICT) is expected to play a key role in this transition. ICT is already a dominant factor for the successful adoption of shared bicycles [2], but we can expect this trend to extend to all others forms of cycling.

Despite the wide consensus about the key role of ICT on new soft mobility paradigms, there are still no clear views on how exactly that potential can be realised. There are major challenges to be addressed, such as limited availability of data sources, their strong dispersion among multiple stakeholders and the lack of clear value propositions to help prioritizing data needs. As a consequence, the lack of systematic and consistent monitoring processes remains one of the obstacles towards more sustainable development of

H. Santos et al. (Eds.): SmartCity 360 2019, LNICST 323, pp. 68–79, 2020.
https://doi.org/10.1007/978-3-030-51005-3_8

cycling mobility policies [3]. This lack of information tools for soft mobility modes is a systemic problem and is also a huge challenge for municipalities and other entities, who need this type of information for their decision-making processes.

1.1 Objectives

In this research, we explore the range of sensors which could provide a systematic source of in-bike sensing data. Our main concern is the instrumentation of bicycles with sensors and the identification of any sort of constraints associated with the data collection process itself.

For this work, only bicycle embedded sensors are being considered, which excludes for example, mobile phones or wearables. Even though we consider any type of bicycle, we will use electric bicycles as our key reference to assess the viability of new sensing possibilities, as they seem to offer the most suitable context for the initial large-scale deployment of this type of embedded sensors. They already have a higher cost, meaning that the added cost of sensors can more easily be diluted within the total cost of the bicycle. Even more importantly, they already have electric power and other electronic equipment, allowing these additional sensors to benefit from the technology already present in the bicycle. Regarding sensor viability, our key assumption is that only low-cost sensors should be considered, and that their total cost should have no significant impact on the production cost of a low-end electric bicycle.

Regarding data needs, we are mainly considering the multi-dimensional characterization of cycle routes. Data collected during a cycle route can offer multiple insights about the route, which can be valuable to the cyclist, but also to all the others cyclists and the city itself. In particular, such data could significantly help cyclists to select their routes. Those decisions can involve complex combinations of criteria and a rich characterization of routes could significantly the decision process. This same information could also be valuable for urban planners and particularly for assessing the cycling potential or the safety risks of existing city streets.

Our research question is about the types of sensors that could be more relevant to integrate in urban bicycles to support this type of systematic and automated collection of data about cycle routes. To seek an answer to this broader question, we need to address two more concrete objectives: The first objective is to identify a set of low-cost sensors that could provide relevant data to monitor phenomena of interest to the characterization of cycle routes. This requires a through exploration of the various types of sensors available and the analysis of their potential to help to categorize cycle routes according to the many criteria that can be used to support route choice. The second objective is to assess the implications of using those sensors in real-world cycling situations. This involves experimenting with the data collection process to access the quality and relevance of the data which can effectively be obtained within the many constraints of real-world cycling situations.

2 Related Work

Cycling is a very personal and circumstantial experience, and cyclists can consider a broad range of factors when selecting the best route to a given destination. Understanding these factors is crucial for route planning, but indirectly it is also a way to understand the vital factors which influence cycling preferences and consequently understand how cities can be enhanced to become more cycling friendly. Previous studies have identified many factors which influence route choice, including dependent attributes (e.g. trip distance, travel time, network characteristics, etc.) and generic trip attributes (e.g. socio-demographic characteristics of the cyclists, trip purpose and weather conditions) [4]. While distance and travel time are regularly mentioned as top factors, cyclists are also particularly sensitive to variables which are related to their perceived risk, such as traffic volume, road types or speed limits. A study in Minneapolis [5] analysed over 1000 rides and concluded that the chosen routes were not always the shortest, most of which included the presence of bicycle paths. A survey in the city of Vancouver, based on paper, web and phone questionnaires, has identified that cyclists prefer to ride away from the noise of car traffic and pollution, by routes with beautiful landscapes separated from traffic, by dirt-free roads and finally by streets without high-speed or heavy-haul vehicles [6]. According to Felix [7], the shortest route is not always the most attractive for cyclists, because there are other relevant factors, such as slope of path, distance or safety perception. However, the decision always depends on the rider, and can be strongly situated. Hochmair defined a set of classes of route decision criteria [8]. The best rated were Time, Safety, Simplicity and Attractiveness. Multiple factors can be associated with each class. Examples include distance, crossings, reduced light traffic, cycle paths, night time brightness, floor conditions and avoid heavy-duty traffic.

Smart phones are increasingly seen as a platform for large-scale data collection [9]. They already integrate a very vast range of sensors, enabling the collection of substantial data about people and their movements. Using the smartphone sensors of urban cyclists, the data transmitted by them can enable the generation of collective knowledge in order to improve the quality of cycling mobility. BeCity [10] is an example of a mobile application that allows all riders to share their tracks and comments, working as a distributed data collection system. It also includes the ability to recommend routes, considering factors such as distance, presence of bike paths and even the attractiveness of those paths. Another example is the BikeNet mobile application, which gathers data about the rides to provide cyclists with a general perspective of their experience and performance. This system is able to obtain information about the environment and the entire experience along the way, such as pollution levels, noise and floor condition [11].

Data gathered from sensors is often used to estimate more subjective factors, such as the quality of the road or the comfort it can provide to cyclists. Verstockt et al. [12] combined data from GPS, accelerometers and web-based geographic API's to classify the terrain type (asphalt, mud, earth, parallel). Biketastic [13] also uses smartphone's sensors to access information about location as well as an accelerometer and a microphone, capable of measuring the state of the floor and the noise level all the way. Cyclists can also provide feedback about the routes using written comments, voice clips or photos. This information can then be used to help to choose routes which are more aligned with user preferences. Aeroflex uses bicycles in an urban environment to make measurements about air quality [14]. The information captured by the system is meant to support the identification of pollution hot spots, e.g. major car intersections or industrial zones, classify streets according to their air quality and determine the exposure level to which cyclists are exposed. In our work, we are exploring which of these many forms of data collection could become a common expectation for urban cycling.

3 Experimental Data Collection

A key part of this research involved the collection and analysis of diverse forms of cycling data to gain a deeper insight about the real-world challenges of systematic cycle-based sensing. The objective is to understand the viability of the sensor deployment on the bicycle as well as the quality of the data generated.

Sensor Selection and Deployment
For this study, we considered a very diverse set of low-cost sensors, which could, as much as possible, address the very broad set of criteria which have been referred in the literature as route choice criteria. Table 1 identifies the selected sensors.

Table 1. Sensor types and corresponding route choice criteria.

#sensors	Sensor type	Route choice criteria
1	GPS	Route trace
2	Accelerometers	Road surface quality, ride smoothness
4	Distance	Number of nearby obstacles, surrounding traffic
1	CO2/VOCs	Exposure to pollution
1	Sound level	Exposure to noise
1	Light level	Sunny/shady routes
1	Environmental	Weather conditions (humidity, temperature, pressure)

To control the system, we used an Arduino UNO unit with a Qwiic card to connect to most sensors and a micro SD card to store the collected data. A LED light was also added to signal when the system is in operation. A 12 V battery was used to power the whole system.

An important element for assessing the viability of the whole approach was to analyse the implications of deploying these sensors in a bicycle for real-world usage. In this work, we only considered the implications of sensor position and connection. We did not consider other issues such as protection against theft, vandalism or exposure to environmental elements. The deployment of the sensors was made easier by the existence of a front basket in the bicycle. The basket provided a natural and valuable context for placing the Arduino Unit and its battery, which were both placed inside the basket. Many sensors were also attached to the basket to facilitate the cable connections with the Arduino unit. In a production bicycle, these devices would have to be embedded within the bicycle itself, which could produce additional deployment constraints.

Fig. 1. Prototype bicycle for data collection

Sensor deployment had to be made according to the particular properties of each sensor. Placement was absolutely critical for some of those sensors, while largely irrelevant for others. In Fig. 1, we can visualize the prototype for data collection, with the various sensors placed at key locations according to their properties.

We had to consider the specific placement requirements associated with distance and light sensors, which were meant to assess proximity to obstacles or moving vehicles or people. They were placed pointing at four distinct directions: forward, right, left and rear. For the placement of the first three, we took advantage of the existence of the bicycle basket, where it was easier to accommodate them all, while complying with directional requirements. The fourth sensor, on the rear, was attached directly to the bicycle frame underneath the seat. The light sensor was also placed facing up in the same basket to maximise exposure to light. It was also placed as far as possible from the cyclist, in this case at the front of the bicycle basket, to reduce the interference of the cyclist's shadow in the collected data. With accelerometers, we considered the need to support slightly different sensing goals. A first sensor was attached to the bicycle frame to maximise the sensing of vibrations. This should be the key data source to assess the smoothness of the road, as well as the use of brakes. A second accelerometer was placed on the handlebar and could serve mainly to assess the smoothness of driving and possibly assess the frequency and intensity of turn movements.

Data Collection Route

To optimize data collection, we defined a specific route, which included streets with various levels of traffic, different road surfaces, areas with natural shade and potentially more polluted areas. The data collection involved 3 rides conducted by two distinct cyclists on three different occasions. As a complement to the sensors, we also used an action camera. This camera was attached to the bicycle and facing forward. Its purpose was to serve as a source of ground truth data to help with the interpretation of the sensor data. Likewise, we also used the MapMyRide mobile application, to register the track and later export it as a GPX file with longitude and latitude data.

Data Analysis

While previous work has already proposed a broad range of data analysis techniques for route characterisation from raw sensor data, the shortage of data remains a major problem for a more systematic use of those techniques. Consequently, our assessment of the data collected was mainly focused on the perspective of its quality and its viability to support common use cases of automated route characterisation.

To support this type of analysis, we developed a set of python scripts to process the data and produce relevant visualizations. For each source, we produced a set of quantitative indicators, mainly in the form of common summary statistics, such as the count, mean, standard deviation, minimum value, maximum value, and quadrant distribution (25%, 50%, 75%). This was complemented with graphic visualizations of the respective time series, which provided important insights into outliers and other abnormal cases. Additionally, we also used the ELAN tool [15] to synchronize the various data sources and link them with the video feed captured by the action camera (c.f. Figure 2).

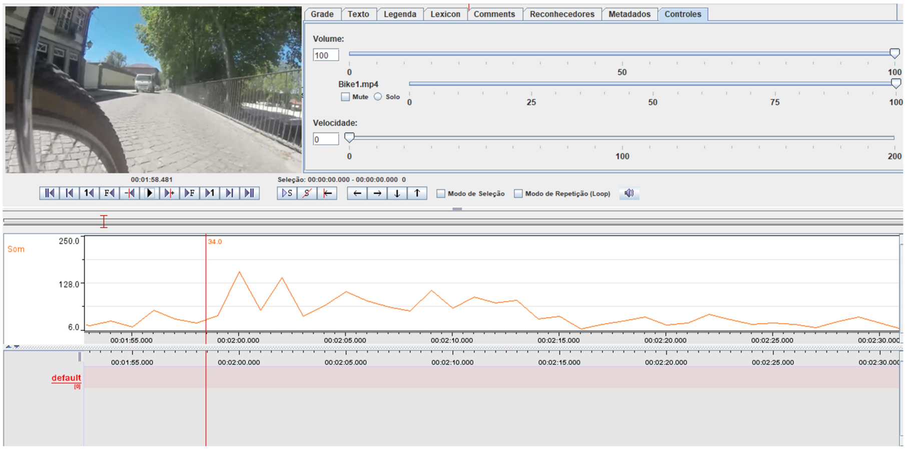

Fig. 2. Elan displaying the video feed and the sound data capture

Even though the video feed was not explored as a data source on its own, it played an important role in this study as a ground truth source. More specifically, it allowed us to explore the concrete situations in which data had been generated and seek to understand the connection between data variations and specific events.

4 Results

The assessment made for the various types of sensors in this study has shown interesting insights about the potential of the various sensors for systematic sensing in a cycling context. The sensors for pressure, humidity and temperature have all behaved as expected, but they did not seem to be the most relevant. Given the relatively low spatial resolution of these data, it can be efficiently obtained through a smaller number of sensors, possibly more reliable ones, at fixed locations around the city.

Regarding accelerometers, we also obtained the type of results which was expectable. This type of sensor is known to be reliable and even low-cost sensors seem to provide a robust data source for most of the more common use cases, e.g. inferring road conditions or level of driving smoothness. We will now focus the analysis on the cases where we were able to uncover more meaningful insights.

4.1 Distance Sensors

In our data collection prototype, we included 4 IR distance sensors, which were supposed to provide data to access the level of obstacles within a short range of the bicycle. These could be buildings, walls, people, parked cars, or traffic in general. The sensors had a 4 meters range and they were placed on the bicycle to cover all the surroundings of the bicycle, at the front, right, left and rear sides. In this case, the data produced has shown multiple incoherences, which might mean that these low-cost sensors were not the most suitable for this task. To better understand the nature of the data produced and the challenges of a correct interpretation of its meaning, we can analyse the graph in Fig. 3, which represents the visualization of the data generated by the front sensor.

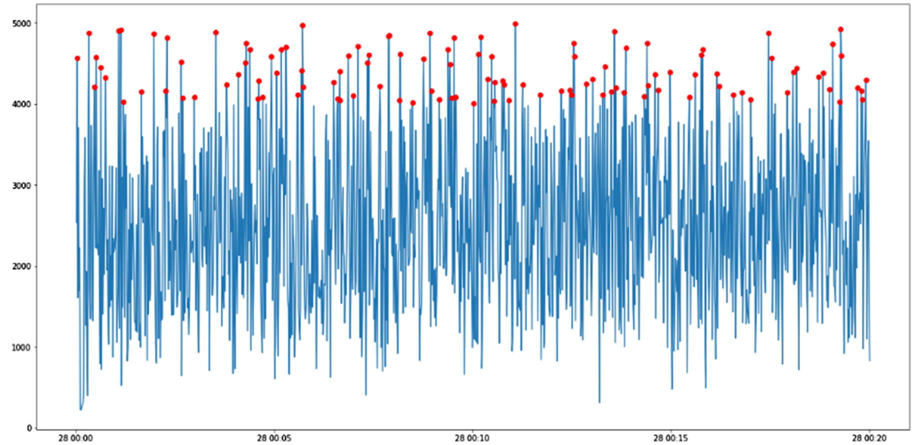

Fig. 3. Distance data produced by the front sensor

The red dots in the graph are signalling the cases where the distance produced exceeded the maximum sensor range of 4 meters (4000 mm). This data seems to suggest the very frequent presence of all sorts of obstacles in front of the bicycle. Data is quickly changing from just about 50 cm to more than 4 meters. However, by comparing this data with the video stream, we were able to confirm that only very rarely there was actually any obstacle within the 4 meters in front of the bicycle. This suggests that the operational requirements for this type of sensing context are not compatible with the type of low-cost distance sensor used in our study. Very similar results were obtained with the other distance sensors, as shown in Fig. 4, which represents right and left sensors.

Once more we can observe an extreme and frequent variation of the data. The first graph represents the data from the right-side sensor and the second from the left side sensor. In this case, the apparently closer distance to objects observed in the right-side sensor seems to be aligned with the intuitive idea that there are normally closer objects

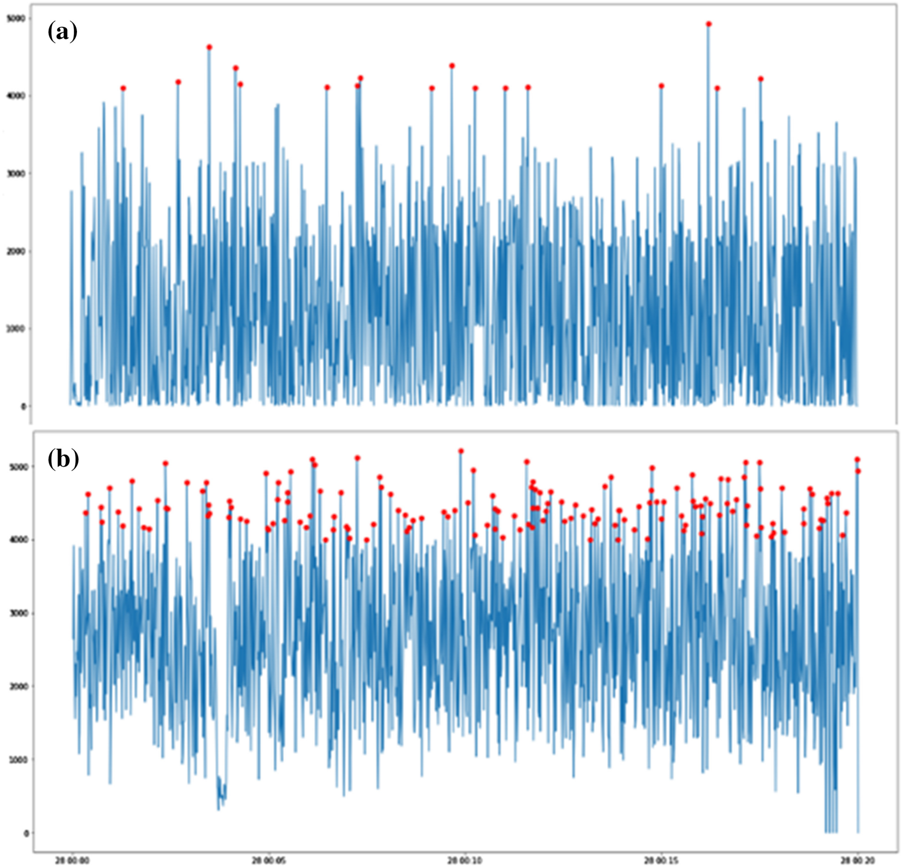

Fig. 4. Distance data produced by right (a) and left (b) sensors (Color figure online)

on that side, such as buildings or parked cars. However, while this was true for some route segments, it was far from being the general rule, as suggested by the graphs. When we used the video to compare the data capture situations, we observed that significant parts of the route were very uncluttered streets without any near obstacles on any side and very low traffic.

4.2 Air Quality Sensors

Air quality sensors can be very relevant to assess the level of exposure to pollutants faced by cyclist. This can help with route selection, especially for commuters, who may take the same cycle route every day and may want to improve the quality of their daily ride. This type of data may also be useful as a new generic source of data about urban pollution. Leveraging bicycles for this particular purpose could be interesting because they are not polluters themselves and collectively they would have the capability to collect large sets of samples every day at multiple location across the city.

The broader concept of air quality can be assessed by measuring some of the specific elements which more commonly correlate with reduced air quality. Carbon Dioxide (CO2) sensors are particularly efficient as a sign of high emissions of pollutant levels and the presence of the others types of pollutants associated with combustion sources, such as traffic or industry. Additionally, Carbon Dioxide sensors are known for being reliable and accurate, even without any type of calibration. In our data collection prototype, we included a carbon dioxide sensor, which measured the concentration of carbon dioxide particles along the selected route, as shown in Fig. 5.

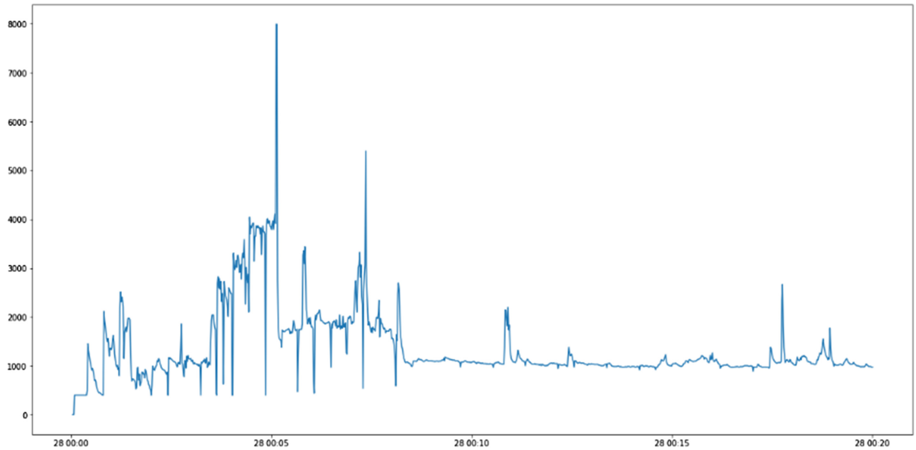

Fig. 5. Data from carbon dioxide sensor

We also included a sensor for Volatile Organic Compounds (VOCs), which could sense a much broader variety of particles. While a VOC sensor may complement Dioxide Carbon sensors, the results we obtained were very similar for both sensors and suggest that a single CO2 sensor may provide a suitable indicator of higher pollution levels.

Given the low-cost nature of these sensors, they should not be seen as a replacement for the more reliable sensors used in official air quality control systems, which many cities have in operation. Still, "peaks" observed in the Fig. 5 were confirmed to correspond to streets where the concentration of pollutants was expected to be higher due to heavy traffic.

4.3 Sound Level Sensors

Figure 6 is a graphical representation of the sound volume level in decibels registered in by the sound sensor during the data collection.

The average sound value was about 46 decibels, a value considered tolerable and normal by human hearing. However, we also identified several outliers were the sound level was above 120 decibels, which even exceeds the human hearing capacity. The video stream allowed us to confirm that no sound of that nature had ever been experienced by riders.

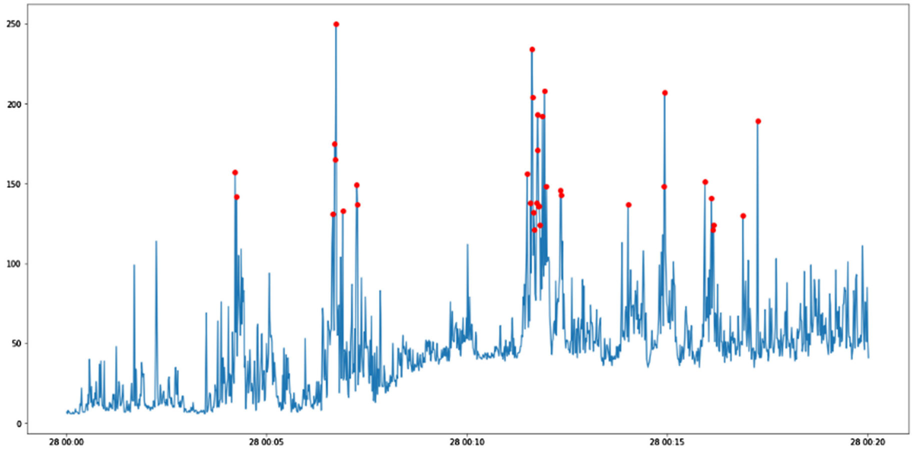

Fig. 6. Data from the sound volume sensor in decibels

4.4 Luminosity

Regarding light sensors, the most striking results is the almost permanent saturation of the data, which seems to indicate that the sensor reached its range limit. Rather than measuring the various levels of light, this data can only allow us to identify occasional situations where the buildings, trees, the cyclist or other sources of shadow covered the sensor. One possible explanation is that this type of low-cost sensor is actually designed to identify these extreme situations of whether or not there is light, rather than accurately measuring the level of light.

5 Discussion

The results have shown some of the practicalities associated with bicycle sensing, allowing us to identify some important design implications for any forms of systematic in-bike sensing.

The first implication to is the huge differences in accuracy, precision and reliability which seem to exist for some sensor types between those normally used for scientific data collection and those which one can reasonably expect to have embedded on any bicycle. In particular, the data produced by a significant set of sensors did not seem to have a plausible alignment with the perception of the concrete situations in which data was captured. Distance sensors seemed to have failed to produce data which is aligned with the reality of the cycle routes in which it captured. They did not seem to be suitable for this role, possibly because their operation properties may not be suitable for the continuously fast changing dynamics around a moving bike. This suggest that a realistic collection of data about the surroundings of a moving bicycle may require much better distance sensors or even more sophisticated techniques, such as Light Detection And Ranging (LiDAR)-based sensor systems [16], similar to those used in drones or autonomous cars. Likewise, sound level sensors have also failed to produce a reliable account of the level of noise experienced by riders during data collection. They have produced unrealistic maximum levels of noise and they have even failed to signal the phases during which the riders were crossing what should be noisier areas. A partial explanation may be associated with the noise produced by the bicycle itself, especially in certain road surfaces. Other sensors, such as light, have given contradictory signs. Even if they have failed to produce an account of the level of light across the route, they were still able to distinguish extreme situations and signal passage through shady areas.

This apparent lack of basic reliability is particularly negative, as it may jeopardize the viability of using some of these sensors. Among the possible reasons is the selection of low-cost sensors, most of which were not designed to be operated under the challenging circumstances of in-bike sensing. They are very sensitive electronic devices, which have been optimised for measurements in controlled and stable situations. They may not be able to cope with a moving bicycle, with its speed, vibrations and quickly changing environment. While this does not necessarily dismiss the use of low-cost sensors, it represents an additional challenge in regard to the cost/benefit equation.

A second implication concerns the positional requirements of some of the sensors. For example, distance sensors can have very stringent requirements about where they should be and where they should be pointing. Small variations in position, may affect the results produced by different bicycles, and the additional complexity brought by this type of requirement, may negatively impact bicycle production costs.

The final implication is to acknowledge that that the range of sensors deployed across bicycles does not need to be uniform. Ultimately, they will be determined by the value propositions that they can offer to cyclists or to a bicycle operator, e.g. in shared bicycles schemes. Still, collectively they should be able to complement each other and benefit from the fact that for some data needs even occasional samples by a reduce subset of the cyclists could still be enough to provide valuable data. Likewise, some key factors can be extrapolated from different types of data. For example, the general riding comfort offered by a particular route may be indirectly estimated from very different sources such as accelerometers, average speed or distance sensors.

6 Conclusions

The contribution of this work is to highlight some of the challenges that can be faced by systematic sensing for urban cycling. Even though previous work has extensively explored the use of various type of data to infer road conditions or traffic situations, this work highlights how the first challenge is actually to produce consistent data in a systematic way. In particular, it seems that the ability to select sensors with operational properties which are suitable for the specificities of their on-bike deployment will play a crucial role in the viability of the whole process. Together with emerging initiatives on standards for cycling data, this type of systematic data collection could significantly impact the ability of cycling ecosystems to really make use of data as a central element for new mobility paradigms. The set of design implications emerging from this work should help to develop new approaches for the systematic integration of sensors in urban bicycles.

Acknowledgements. This work has been supported by national funds through FCT, Fundação para a Ciência e Tecnologia, within the Project Scope: UID/CEC/00319/2019, and also by the European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project nº 039334; Funding Reference: POCI-01-0247-FEDER-039334].

Development of the Arduino code was made by André Torrinha, Marcelo Alves, Pedro Lobo and Rui Almeida as part of a Masters' Project.

References

1. Violeta Bulc: Cycling: green and efficient transport for the future. European Commission (2016). https://ec.europa.eu/commission/commissioners/2014-2019/bulc/blog/cycling-green-and-efficient-transport-future_en. Accessed 31 Oct 2019
2. Lee, J., Leem, Y.T., Lee, S.H.: Categorizing U-Bike service and assessing its adoptability under it-. In: 12th World Conference for Transportation Research, pp. 1–10 (2010)
3. Ricci, M.: Bike sharing: a review of evidence on impacts and processes of implementation and operation. Res. Transp. Bus. Manag. **15**, 28–38 (2015)
4. Li, S.: Cycling in Toronto: route choice behavior and implications to infrastructure planning. Master thesis. University of Waterloo (2017)
5. Harvey, F., Krizek, K.: Commuter bicyclist behavior and facility disruption. Transp. Res. Board **60** (2007)
6. Winters, M., Davidson, G., Kao, D., Teschke, K.: Motivators and deterrents of bicycling: comparing influences on decisions to ride. Transportation **38**(1), 153–168 (2011)
7. Félix, R.: Gestão da Mobilidade em Bicicleta Master thesis, Engenharia do Território. Universidade Técnica de Lisboa (2012)
8. Hochmair, H.: Decision support for bicycle route planning in urban environments. In: 7th AGILE Conference on Geographic Information Science, pp. 697–706 (2004)
9. Srivastava, M., Abdelzaher, T., Szymanski, B.: Human-centric sensing. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **370**(1958), 176–197 (2012)
10. Torres, S., Lalanne, F., Del Canto, G., Morales, F., Bustos-Jimenez, J., Reyes, P.: BeCity: sensing and sensibility on urban cycling for smarter cities. In: Proceedings - International Conf of the Chilean Computer Science Society, SCCC 2016 (2016)
11. Eisenman, S.B., Miluzzo, E., Lane, N.D., Peterson, R.A., Ahn, G.-S., Campbell, A.T.: BikeNet. ACM Trans. Sens. Netw. **6**(1), 1–39 (2009)
12. Verstockt, S., Slavkovikj, V., De Potter, P., Van De Walle, R.: Collaborative bike sensing for automatic geographic enrichment: geoannotation of road/terrain type by multimodal bike sensing. IEEE Signal Process. Mag. **31**(5), 101–111 (2014)
13. Reddy, S., Shilton, K., Denisov, G.: Biketastic: sensing and mapping for better biking. In: Proceedings of the 28th International Conference on Human factors in Computing Systems, pp. 9–12 (2010)
14. Elen, B., et al.: The Aeroflex: a bicycle for mobile air quality measurements. Sens. (Switz.) **13**(1), 221–240 (2013)
15. Auer, E., et al.: ELAN as flexible annotation framework for sound and image processing detectors. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010), European Language Resources Association (ELRA), pp. 890–893 (2010)
16. Zhao, M., Stasinopoulos, S., Yu, Y.: Obstacle detection and avoidance for autonomous bicycles. In: IEEE International Conference on Automation Science and Engineering, 2017 August, pp. 1310–1315.1 (2018)