Crowdsourcing to Mobile Users: A Study of the Role of Platforms and Tasks

Vincenzo Della Mea   Eddy Maddalena   Stefano Mizzaro
Department of Mathematics and Computer Science
University of Udine
Udine, Italy
vincenzo.dellamea@uniud.it, eddy.maddalena@uniud.it, mizzaro@uniud.it

ABSTRACT

We study whether the task currently proposed on crowdsourcing platforms are adequate to mobile devices. We aim at understanding both (i) which crowdsourcing platforms, among the existing ones, are more adequate to mobile devices, and (ii) which kinds of tasks are more adequate to mobile devices. Results of a user study hint that: some crowdsourcing platforms seem more adequate to mobile devices than others; some inadequacy issues seem rather superficial and can be resolved by a better task design; some kinds of tasks are more adequate than others; and there might be some unexpected opportunities with mobile devices.

Categories and Subject Descriptors

H.4.m [**Information systems applications**]: Miscellaneous

General Terms

Experimentation, Measurement.

Keywords

Crowdsourcing, mobile devices.

1. INTRODUCTION AND AIMS

Among the phenomena that are acquiring increasing importance in the information technology landscape, two are the subjects of this paper: (i) crowdsourcing, and (ii) mobile devices and applications.

Crowdsourcing, i.e., the outsourcing of tasks typically performed by a few experts to a large crowd as an open call, has been shown to be reasonably effective in many cases, like Wikipedia, the Chess match of Kasparov against the world in 1999, and several others (see, e.g., [4] or even `http://en.wikipedia.org/wiki/Crowdsourcing`). Several *crowdsourcing platforms* (Amazon Mechanical Turk being probably the most known) have also appeared on the Web: they allow requesters to post the tasks they want to crowdsource and workers to perform those tasks for a small reward (usually a few cents).

Meanwhile, mobile devices (phones, smartphones, tablets, and in the near future glasses, watches, and so on) have become ubiquitous and are used to access the Web. According to several statistics, in the next few years there will be more Web accesses by mobile devices than by classical desktop/laptop computers (see, e.g., [6]).

In this paper we study the intersection of mobile and crowdsourcing. We aim at understanding whether the task currently proposed on crowdsourcing platforms are adequate to mobile devices. By "adequate" we mean that they can be performed effectively by using a mobile device in place of a desktop/laptop computer. We specifically seek to answer two research questions:

Q1 Which crowdsourcing platforms, among the existing ones, are more adequate to mobile devices?

Q2 Which kinds of tasks are more adequate to mobile devices?

Besides the above mentioned statistics on increasing mobile usage, this research is also justified by the fact that today quite often people access the Web on their mobile phones for short periods of time, for example while commuting to work on train or underground, while waiting for a bus or for a friend, while in a car (and not driving), while standing in a queue, etc. In other terms, there is plenty of human workforce available for a few minutes (or seconds) bursts, and this kind of workforce seems perfect for the crowdsourcing scenario, where the tasks are usually short and the reward is usually low. Moreover, some crowdsourcing tasks could be more adequate to a mobile scenario than to a classical desktop one. For example, taking pictures of some point of interest (like a monument, a paint, or a billboard), describing a real life scene, or even recording movements, destinations, and trajectories in an urban traffic setting. However, to fruitfully exploit this workforce, it is necessary that the platforms are adequate and tasks are feasible. This consideration also underlies our choice of focussing on the worker side and neglecting the requester part.

The paper is structured as follows. In Section 2 we briefly survey the related work on mobile and crowdsourcing, trying to focus on the research involving both aspects. In Sections 3 and 4 we describe two experiments aiming at answering the two research questions above. In Section 5 we draw conclusions and sketch future developments.

2. RELATED WORK

Although crowdsourcing commercial platforms seem designed with a desktop/laptop user in mind, there has already been some work on the idea of having workers using mobile devices. We briefly survey it in this section.

Musthag and Ganesan[7] focus on mobile micro-task market and present some statistics on mobile workers behavior.

The mCrowd platform [11] is an iPhone based mobile crowdsourcing platform that enables mobile users to act as both requester and workers, and focuses on tasks like geolocation-aware image collection, road traffic monitoring, etc., that exploit the rich array of sensors available on iPhones.

Eagle [2] describes txteagle, a mobile crowdsourcing marketplace used in Kenya and Rwanda for tasks like translations, polls, and transcriptions.

Location-based distribution of tasks to mobile workers is proposed in [1]. Some design criteria for mobile crowdsourcing platforms are also presented and discussed. A similar approach, focused on the specific domain of news reporting is presented in [9]: SMS messages are used for location based assignment for crowdsourcing news.

Narula and colleagues [8] focus on low-end mobile devices and present MobileWorks, a platform for OCR tasks specifically aimed at users from the developing world. Experimental results demonstrate a high rate of task completion (120 per hour) and a high accuracy (99%). A similar approach is presented in [3], where the mClerk system is described. Some experimental results again witness the feasibility of the approach. Some discussion of the viral diffusion of the system among workers is also discussed.

As a different approach, the CrowdSearch system, an image search service for mobile phones that relies on Amazon Mechanical Turk, is presented in [10]. It is interesting because, although it does not exploit a mobile crowd, it is an example of exploiting a crowd in (almost) real time.

3. EXPERIMENT 1

3.1 Aims

The first experiment aims to verify the suitability of existing crowdsourcing platforms to mobile devices (see question Q1 in Section 1). We asked the participants to estimate the difficulty of performing a task on both a mobile device and a desktop/laptop computer.

3.2 Participants

Sixteen participants were involved in the experiment. All of them were italian students, aged between 16 and 30. We required a good knowledge of English and familiarity with computers and smartphones. Participants were randomly subdivided into 4 groups (U_1,U_2,U_3,U_4), each one containing four participants.

3.3 Data

We selected four among the most popular crowdsourcing platforms (see Table 1). We downloaded some randomly selected tasks from these platform, for a total of 2717 tasks (the exact number for each platform is shown in the third column in Table 1). The download has been performed in October and November 2012. The downloaded tasks are among those that can be performed by any requester, i.e., without any qualification. These are not huge samples: for example, on mTurk one can count hundreds of thousands of tasks available per month [5]. Though, the samples are neither negligible, since they count around 1% − 5%. For each task we extracted: identifier, title, required proof, remuneration, time needed, requester identifier, and description. The task collection is available upon request. Three examples of tasks in our collection are (errors included):

id	Platform name	URL	# of tasks
mTurk	Amazon Mechanical Turk	`mturk.com`	1154
micW	Micro Workers	`microworkers.com`	1302
minW	Minute Workers	`minuteworkers.com`	86
shortT	Short Task	`shorttask.com`	175

Table 1: Platforms

- **Task example 1:**
  1. *Go to http://goo.gl/Dlzk*
  2. *Click the link to go to the download*
  3. *Complete a survey/offer on Sharecash and download the file*
  4. *Send proof*
- **Task example 2:**
  1. *Go to http://OneDollarRiches.com/5737*
  2. *Click on Join Now button*
  3. *Invest 1 dollar by logging in into your Alertpay account*
  4. *After that enter you personal details and login.*
  5. *Join and finish signing up*

  While Sign up use same e-mail of your Alertpay account. because when u make ur refferaf there 1$ sing up go direct into ur alterpay account.
- **Task example 3:** *Find the details for this Restaurant*
  – *For this restaurant below, enter the details below*
  – *You must confirm that the restaurant is still open*
  – *Include the full address, e.g. http://www.thecheesecakefactory.com*
  – *Do not include URLs to city guides and listings like Citysearch*

  Restaurant : Akasha Organics 160 North Main St. Ketchum
  Fill in the text fields with this information: Still open, Restaurant name,Website Address,Phone number,Street Address,City,State,Zip code.

3.4 Methods

We randomly extracted 48 tasks, 12 from each platform, and divided them into 4 groups (T_1, T_2, T_3, T_4). Each group contains 12 tasks (3 tasks from each of the 4 platforms). Task group T_i was assigned to user group U_i (e.g., task group T_1 was assigned user group U_1). We developed a web application to show to each participant the group of 12 tasks assigned to his/her user group (see Figure 1). By using this

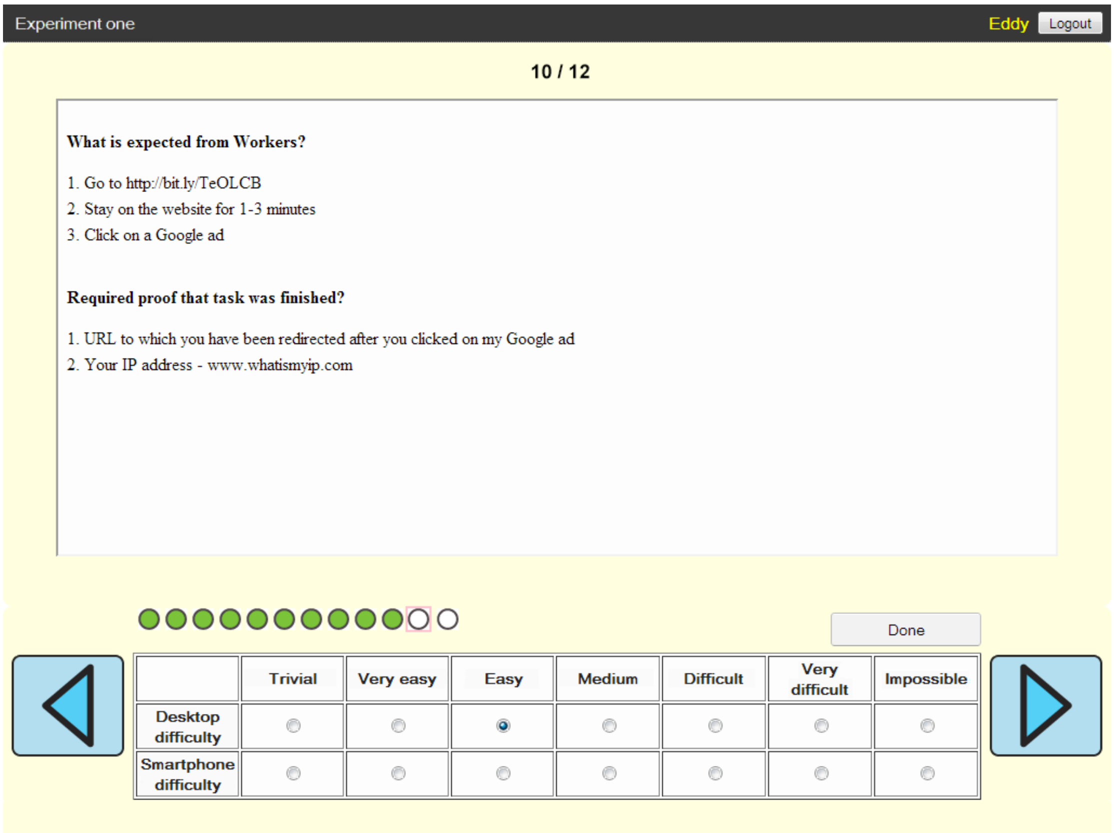

Figure 1: The interface used in the first experiment (translated into English)

application, each participant recorded two estimates of difficulty for each task, one for a desktop and one for a mobile device (see the bottom part of the figure). Tasks were presented in random order and participants did not know from which platform the tasks were extracted.

Difficulty was provided on a seven points scale ranging from trivial to impossible. For each task we therefore obtained 4 estimates (from the participants in the same group). We then converted the labels into the [0..6] range and calculated the average of difficulty estimates.

3.5 Results

Figure 2 shows the averaged estimated difficulty, on desktop and mobile, for each platform. Tasks from mTurk are estimated slightly more difficult than MicroWorkers, MinuteWorkers, and ShortTask. The difference of difficulty estimates between desktop and mobile is also shown in Figure 3: difficulty estimation is consistently higher on mobile devices, both in absolute terms and as a percentage of the desktop difficulty.

By manually analyzing the task collection we realized that some of them are inadequate to mobile devices for some typical reasons:

- too long description;
- technical obstacles like scrolling problems, unsupported audio formats and/or plugins, pages with Adobe Flash, etc.;

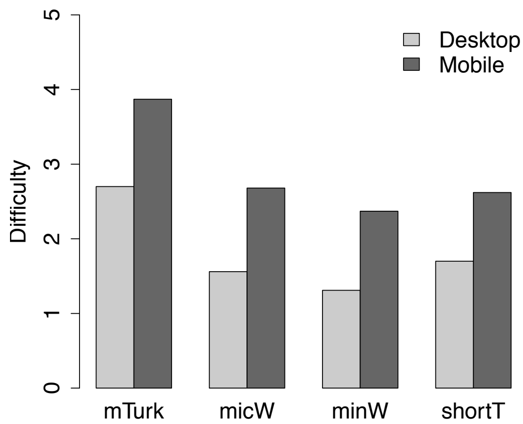

Figure 2: Estimated difficulty

- use of frame attribute in html pages;
- bad layout in a small resolution display;
- need of a high power CPU.

Some of these task issues seem due to the task content, while some other depend on how the Web interface is realized. Many of them seem rather superficial and can be overcome by a better task design and/or better user interfaces.

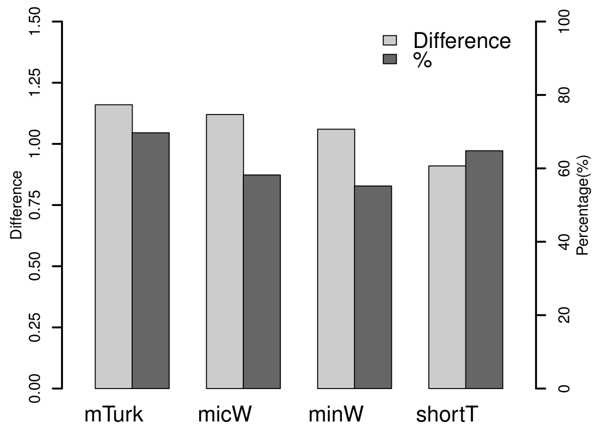

Figure 3: Mobile-desktop difference of estimated difficulty, as absolute time (bars on the left) and as a fraction (right)

4. EXPERIMENT 2

4.1 Aims

The aim of the second experiment is to identify which task kinds are more adequate for mobile devices (see question Q2 in Section 1). We therefore now focus on task features, and not on platforms. Also, in place of asking estimates to participants, we required them to actually perform the tasks on both desktop and mobile devices and we measured the time spent on each task. Participants used two prototype platforms that we built ad hoc for the experiment: one for desktop devices using Google Web Toolkit, and the other specifically made for mobile devices, by means of an Android application. Figure 4 shows the resulting user interfaces.

4.2 Participants and Data

The 16 participants (the same as in the previous experiment) were subdivided into 4 groups labeled U_1, U_2, U_3, U_4.

To identify the kinds of task in a somehow objective way, we relied on the task categories usually requested in crowdsourcing marketplaces. More in detail, we started from the 11 categories suggested by Amazon Mechanical Turk when creating a new task (see `https://requester.mturk.com/create/projects/new`): Categorization, Data Collection, Moderation of an Image, Sentiment, Survey, Survey Link, Tagging of an Image, Transcription from A/V, Transcription from an Image, Writing, and Other. To obtain an amenable number of categories in our experiment, we excluded 5 Mechanical Turk categories: Data collection, Survey and Survey link (considered somehow similar to Sentiment), Transcription from A/V (to avoid technical issues on mobile devices), and Other. We therefore selected 6 task categories, those shown in Table 2. Then we created 4 new tasks for each category, for a total of 24 tasks, and grouped them in four task groups (labeled T_a, T_b, T_c, T_d), each group containing six tasks, one from each category.

Using artificial tasks (i.e., tasks created by ourselves) allowed to remove any platform bias and those issues discussed at the end of Section 3.5, that might have affected the results. Also, their classification was easier (sometimes it is not clear how to classify real tasks). Finally, this allowed us to create task descriptions written in Italian, thus removing any language issue from the experiment (all participants were Italian native speakers). The created tasks are in all respects similar to real tasks.

4.3 Methods

We took the usual special care to avoid any order and learning bias. Each participant performed 6 tasks (one for each of the categories in Table 2) on the desktop platform and 6 other tasks (again, one for each category) on the mobile one. His/her tasks were selected from two task groups, depending on the user group the participant was assigned to. To further avoid bias, participants in each group alternatively started from desktop or from mobile. Therefore, each participant performed a total of 12 different tasks, half on desktop and half on mobile. Each task was performed by 8 participants in two user groups, half of which performed it on mobile and half on desktop.

Statistics have been calculated as follows. At first, the average time needed for task completion has been calculated for each task separately for mobile and desktop performance (i.e., averaged on 4 subjects each). Then category averages have been calculated from task averages, again separately for mobile and desktop devices.

4.4 Results

Figure 5 shows the average time to complete for a task, for each category and on both mobile and desktop devices. Figure 6 shows the differences in average time to complete. Some tasks are quicker: Cat, Mod, Sen required less than one minute on average, on both desktop and mobile. ImT and Tra are a bit longer, between one and two minutes on average, and Wri is even longer. As expected, all tasks are faster on desktop, with the only exception of Wri: in it, the participants autonomously decided to use the voice-to-text functionality when on mobile, and this turned out to be quicker than writing with a keyboard (although we did not investigate the quality of transcription). As highlighted in Figure 6, ImT and Tra show a higher mobile-desktop difference, both on absolute time and percentage, probably because they require multiple texts in more fields, a cumbersome activity if carried out by mobile.

Looking at the percentage differences in Figure 6, one can notice that Cat small difference in absolute terms is actually quite high in percentage: this means that even if the difference in time is rather small, since Cat tasks are quite short (as can be seen in Figure 5), this small value is important in percentage terms. Conversely, looking at the two rightmost bars, the percentage difference in Wri looks smaller than the absolute time difference; this is again due to the average length of the Wri task, which is quite high (see Figure 5). Though, the improvement on mobile is still important, being around 20%.

5. CONCLUSIONS AND FUTURE WORK

The work described in this paper is a first exploration of the opportunities and challenges of outsourcing tasks to a mobile crowd. Results provide preliminary evidence on the inadequacy of current crowdsourcing platforms for mobile devices, even if task complexity would be adequate for being

Figure 4: The interface used in the second experiment: desktop (left) and mobile (right)

Id	Category	Description
Cat	Content categorization	Some images are proposed to the worker, which is required to assign each of them to the correct category.
Mod	Moderation of an image	The worker is required to flag adult contente pictures that are inappropriate for children.
Sen	Sentiment	Some sentences are proposed to the worker, which is required to record his agreement by means of a Likert scale.
ImT	Image tagging	Some images are proposed to worker, which is required to tag each of them with keywords.
Tra	Transcription from an image	The worker is required to extract and write the textual content from a picture.
Wri	Writing	The worker is required to write a short text about a specific topic.

Table 2: Task categories

carried out on mobile scenarios. More in detail, results are fourfold:

- Experiment 1 results show that, according to user perception of difficulty, some crowdsourcing platforms might be slightly more adequate to mobile devices than others.

- Some inadequacy issues seem rather superficial and can be resolved by a better task or interface design.

- Experiment 2 shows that tasks of different kinds, as defined by mTurk categories, might present different difficulties when carried out on desktop or on mobile devices. This might hint a first specialization of task assignment, although examining features of easy and difficult tasks might provide a better ad-hoc specialization, perhaps even independent of the kind of task.

- Experiment 2 also confirms that mobile devices might offer some unexpected opportunities, like the voice-to-text unexpected (by us) solution, autonomously adopted by participants.

We carried out two separate experiments, although sharing subjects, in order to study two different aspects of mobile crowdsourcing: crowdsourcing platform effects, and task category effects. The experiments are preliminary and results are not final, but this is consistent with our aims, that were to begin to study the general issue of mobile crowdsourcing. This exploratory attitude is also a motivation for having two experiments performed with different methodologies (asking to the participants an estimate of difficulty and having participants performing the actual tasks). Of course, these experiments, or similar ones, could have been run by means of some crowdsourcing platform themselves. We preferred a more traditional approach and started with

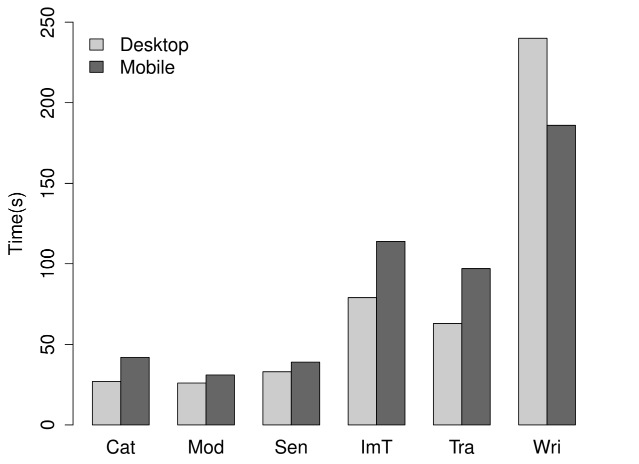

Figure 5: Average time to complete for each task category on both mobile and desktop devices

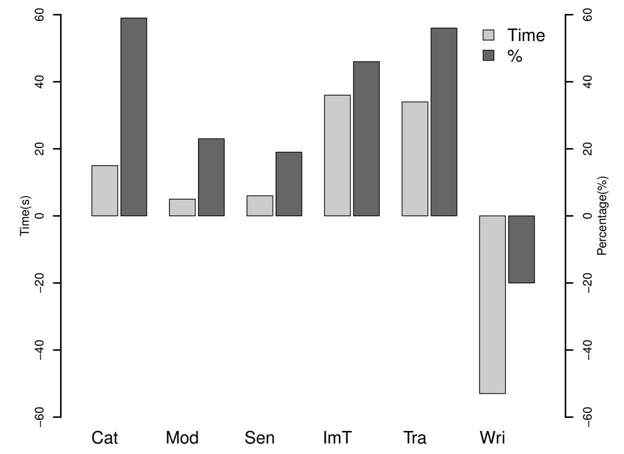

Figure 6: Mobile-desktop differences in average time to complete for each task category, as absolute time (bars on the left) and as a fraction (right)

classical user studies, but we do plan to do that in the future.

To further develop this work, other experiments can be imagined. For example, the same experiments described here could be repeated in real-world scenarios (on the train, road, school rooms, or crowded places) to have more realistic results. It is also feasible to imagine an extended crowdsourcing platform that on the basis of the context of a worker (time, date, geolocation, habits and preferences, mobile device sensors, etc.), automatically filters and selects tasks tailored for a specific context.

6. REFERENCES

[1] F. Alt, A. S. Shirazi, A. Schmidt, U. Kramer, and Z. Nawaz. Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 13–22, New York, NY, USA, 2010. ACM.

[2] N. Eagle. txteagle: Mobile crowdsourcing. In *Proceedings of the 3rd International Conference on Internationalization, Design and Global Development: Held as Part of HCI International 2009*, IDGD '09, pages 447–456, Berlin, Heidelberg, 2009. Springer-Verlag.

[3] A. Gupta, W. Thies, E. Cutrell, and R. Balakrishnan. mClerk: enabling mobile crowdsourcing in developing regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1843–1852, New York, NY, USA, 2012. ACM.

[4] J. Howe. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Random House Inc, 2008.

[5] P. G. Ipeirotis. Analyzing the amazon mechanical turk marketplace. *XRDS*, 17(2):16–21, Dec. 2010.

[6] M. Meeker and L. Wu. Internet Trends D11 Conference — The annual Internet Trends Report, 2013. `http://www.slideshare.net/kleinerperkins/kpcb-internet-trends-2013`.

[7] M. Musthag and D. Ganesan. Labor dynamics in a mobile micro-task market. In W. E. Mackay, S. A. Brewster, and S. Bødker, editors, *CHI*, pages 641–650. ACM, 2013.

[8] P. Narula, P. Gutheim, D. Rolnitzky, A. Kulkarni, and B. Hartmann. MobileWorks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. *Proc. HCOMP11*, 2011.

[9] H. Väätäjä, T. Vainio, E. Sirkkunen, and K. Salo. Crowdsourced news reporting: supporting news content creation with mobile phones. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 435–444, New York, NY, USA, 2011. ACM.

[10] T. Yan, V. Kumar, and D. Ganesan. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In *MobiSys '10: Proceedings of the 8th international conference on Mobile systems, applications and services*, pages 77–90. ACM Press, 2010.

[11] T. Yan, M. Marzilli, R. Holmes, D. Ganesan, and M. Corner. mCrowd: a platform for mobile crowdsourcing. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys '09, pages 347–348, New York, NY, USA, 2009. ACM.