



Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application

Benjamin Poppinga, Martin Pielot, Niels Henze, and Susanne Boll

Wi: Journal of Mobile Media, Spring 2011

The online version of this article can be found at:

<https://doi.org/10.65968/TCOA5526>

Poppinga, Benjamin, Martin Pielot, Niels Henze, and Susanne Boll. "Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application". *Wi: Journal of Mobile Media*. Spring (2011). Web. <https://doi.org/10.65968/TCOA5526>

Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application

Benjamin Poppinga, Martin Pielot, Niels Henze, and Susanne Boll

Introduction

To observe the mobile user experience various observation techniques exist. For field studies ethnographic observation techniques, like shadowing, are often used. In shadowing, an experimenter follows a participant and takes notes on the observed behaviour. Shadowing is known to be highly situated [3, 5]. However, this technique does not scale very well. Additionally, because of its obtrusiveness, it could change the observed participant's behaviour.

To overcome the disadvantages of low scalability and high obtrusiveness, new observation methods are being developed. In theory, passive automated logging through sensors seems to reach the same "situatedness", while being scalable and unobtrusive [3, 5]. In practice logging has rarely been applied to mobile observation during the last years. One reason for this might be that suitable data sources, e.g. sensors, were not available on a common mobile device. However, the extension of smart phones through external sensors showed that sensors are able to infer users' everyday situations [2].

Nowadays a commercial mobile smart phone, like the iPhone, has a variety of sensors integrated. Thus, principles where earlier specialized hardware was required, can now be ported to the phone (e.g. a pedometer). McMillan et. al. [4] successfully applied logging in the large scale in a mobile game, which they submitted into the App Store. Hence, these sensors make logging more and more interesting as scalable, unobtrusive, and situated observation technique.

However, while there are some well-known concepts, like a pedometer algorithm available and ready for instant application, a holistic view on how to use, combine, and apply sensors to log a specific user action is missing. In this paper we present our approach towards unsupervised in-market studies and identify three major challenges based on our preliminary findings.



Figure 1: The PocketNavigator is a mobile pedestrian navigation application. Our integrated sensor-based observation technique is invisible to the user. However, the participation within the user study is defined as opt-in to maintain ethical correctness.

Experiment Design

Originating from the interest in providing tactile feedback as an additional navigation aid, we developed the PocketNavigator. The PocketNavigator is a personal navigation application, available for free in the Android market (see Figure 1). It is designed as a traditional map-based application, providing a map surface, the user's location, and a waypoint-based route towards an arbitrary destination [6].

In addition, the application is complemented by a concept, which encodes the direction towards the next waypoint in vibration patterns. If the waypoint is straight ahead of the user, two vibration pulses of equal length are shown. If the next waypoint is on the right, the duration of the second pulse increases. The same happens to the first pulse, if the waypoint is on the left. If the waypoint is behind a user, three pulses are shown.

The additional values we assumed for the tactile feedback are that a user will need to watch the display less often, will commit fewer navigation errors, and will be less disoriented. These three assumptions serve as hypotheses for an experiment we decided to conduct remotely and unsupervised in the Android Market. If a concrete research question should be answered, it is recommended to define the hypothesis right before any sensor data is gathered.

Then, for each hypothesis the observable values need to be identified. Therefore one should think about what are observable events, supporting or not supporting the hypothesis. Comparable studies in literature already propose a definition for how a specific parameter can be observed. In case of the PocketNavigator, we decided to measure e.g. if the user looks at the display by using the roll and pitch angle, determined through the accelerometer.

In the last step the to-be-measured values will be assigned and represented through available sensors. In the exemplary case, if the user is watching the display we decided to use the accelerometer, which is able to provide the required values roll and pitch. As one can imagine, every matching of a hypothesis to an observable behaviour and then to a set of sensors induces some noise and inaccuracy. Thus, it is necessary to design and validate the sufficient representation of a to-be-observed behaviour iteratively. At some

time if the selected representations are reasonably accurate, the experiment can be released to the market.

Identified Challenges

The PocketNavigator is still available and the study (i.e., the logging) is still ongoing. At this point, more than 500 people have participated in the study. In this section we transfer our experiences into general challenges, which need to be approached to further establish sensor-based observation in mobile applications. We identified three challenges: recruiting, analysis, and the question of internal validity.

Recruiting

In the participant recruitment process, a good application title and description needs to be provided to attract participants. Further, a well-designed application icon and some screenshots could also attract users. Without question, the application should provide the advertised functionality, be robust and reliable.

To fulfill ethical requirements, the study must be announced to the user in a clear and transparent way. Thus, the mentioning of the study in the application's general terms and conditions will not suffice. A separate menu entry should be made available clarifying the purpose and frame of the study. The participation in the user study should

be opt-in instead of opt-out. A user should be able to withdraw at any time, at least by uninstalling the application.

Early releases of the PocketNavigator presented the study in a separate info view, selectable through the application's menu. If interested in participation, the user must explicitly check a checkbox. However, under this condition the acquisition of participants proceeded quite slowly. In an updated version, we proactively announced the study through a simple and short pop up dialog. If the user declines to participate in the study, a more detailed info screen on the study is shown, trying to convince the user. This approach has lead to an increased participation rate of about 5 to 10%.

Data Analysis

The recording of sensor values within the application is one thing. However, the gathered data of each client must be available to do analysis. Therefore we used a custom made server where each client connects via sockets and transmits the gathered data in chunks. Alternatively a script, running on an existing server can be used, such as PHP. This can also be easily combined with encryption algorithms, like SSL. To avoid loss of any data, a backup and watchdog is recommended.

Once the application is in the market and the participants are sending their data, one can begin the analysis. From our personal experience, we recommend doing the analysis on a regular basis, to identify overlooked aspects or strange application behaviours,

which can be solved by adapting the logging algorithms. With every adoption it is important to monitor the version a participant is using and to not confuse different types of data during analysis.

The actual analysis is done by custom made tools, as universal analysis tools probably do not exist for a specific use case. In case of the PocketNavigator we build one application, which does a summary over the data of all participants and prepares an output file, which is readable by e.g. Microsoft Excel. Second, we have built an application, which is able to replay the behaviour of an individual user by displaying the values of the sensors in real time. The first tool is more suited for quantitative analysis, while the second tool can give insights in individual's situations, which can be treated as qualitative data.

Internal Validity

In controlled experiments, internal and external validity are two contrasting aims. Internal validity is the validity of the inference of causal relationships, or how confident the observed effects can be attributed to the experimental manipulation. External validity is the validity of the generalization of experimental findings, or how confident the observed findings can be generalized beyond the experiments setting.

Typically, experiments (especially those conducted in the lab) focus on internal validity. The disadvantage of this approach is that the experimenters often generalize their findings to actual usage scenarios. Studying applications in "real" use by making them

available to a wide range of users -as we did with the PocketNavigator -stresses external validity at the expense of the internal validity.

In the case of the PocketNavigator we identified two factors threatening the internal validity: the design as quasi-experiment and the unpredictable usage.

Experiment vs. Quasi-Experiment

In a true experiment, conditions get allocated randomly. As we are studying the effect of the vibro-tactile feedback technique, in a true experiment, half of the participants would be chosen to use the tactile feedback and the other half not. However, in our actual study design, we allowed the participants to choose for themselves if the tactile feedback should be turned on or off. We were afraid that people would get annoyed by the tactile feedback, giving the application bad ratings in the Android Market, and in consequence deterring potential future users.

Thus, the experiment is not a true, but a quasi experiment. Due to the lack of randomization it is harder to rule out confounding variables and unsystematic variance. In our study, people that decide to use the tactile feedback could have certain traits or be in certain situations which favour or disfavour the usage. For example, if only people with lots of experience use the tactile feedback, because they are more open to new innovations, their navigation performance could be disproportionately better than average because of either their experience or the tactile feedback.

Unpredictable Usage

Another problem that turned up is the unpredictable usage of the application. In a typical experiment the task is well defined and well known to the person analyzing the data. In the case of the PocketNavigator, we neither have a way to dictate a certain usage pattern to the users nor can we completely understand the usage at a certain time. In what follows, we will provide a few examples of unpredicted usage patterns that could have threatened the internal validity if we had not identified them:

Example 1: Lying on table. In the first stream of data we received from our participants, we had many situations where no navigation took place. Having a close look at the data, the accelerometer indicated that the device was oriented parallel to the surface and the GPS signal showed no walking speed. From this data, we inferred that many users might be testing the application indoors, possibly leaving the device on the table and probably keep running the application in the background.

Example 2: Car Driving. At a later stage we were investigating the effects of the tactile feedback on the average walking speed. However, we were surprised by the huge variance in the walking speed averages. Taking a closer look at the individual data we found that some walking speeds were unnaturally high (e.g. > 70km/h in average) for pedestrians, hence, we inferred that people had used it in their cars or another moving vehicle.

Example 3: Background idling. Android offers parallel and background executing. As the PocketNavigator is expected to run in the pocket, we designed it to continue running when the screen saver is activated or another application is pushed to the front. The problem is that the Android OS does not really terminate applications but only pushes them into the background until the resources are needed otherwise. Thus, in a few cases the application kept running in the background producing nonsense data.

Conclusion

In this paper we report on our experiences, applying a sensor-based virtual observer to the Android Market. We identify three major issues that should be considered in future developments: recruitment, data analysis, and internal validity.

In our future work, we want to extend and apply the in-market observation methodology for true experiments, as well as for more open research questions, which cannot be answered within an experiment. Additionally, we want to apply logging as an observation method in a traditional field study to prove the validity of the method. Finally we are interested in the advantages, disadvantages, and limitations of the virtual observer in different settings.

Acknowledgments

The authors are grateful to the European Commission which co-funds the IP HaptiMap (FP7-ICT-224675). We like to thank our colleagues for sharing their ideas with us.

References

Choudhury, T., et al. (2008). The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2), 32–41.

Consolvo, S., et al. (2008). Activity sensing in the wild: a field trial of ubifit garden. *In Proc. of CHI*.

Froehlich, J. et al. (2007). My experience: a system for in situ tracing and capturing of user feedback on mobile phones. *In Proc. of MobiSys*. 57–70.

McMillan, D. (2010). Further into the wild: Running worldwide trials of mobile systems. *In Proc. of Pervasive*, 6030, 210–227.

Mulder, I. et al. (2005). SocioXensor: Measuring user behaviour and user experience in context with mobile devices. *In International Conference on Methods and Techniques in Behavioural Research*.

Pielot, M. et al. (2010). Pocketnavigator: Vibro-tactile waypoint navigation for everyday mobile devices. *In Proc. of MobileHCI*.

Biography

Benjamin Poppinga is a scientific researcher in the Intelligent User Interfaces Group at the OFFIS – Institute for Information Technology, Germany. He is also pursuing his PhD studies in the Media Informatics and Multimedia Systems Group at the University of Oldenburg, Germany. He graduated in 2008 in Computer Science from the University of Oldenburg. Benjamin is working in the European research project HaptiMap which aims at making geospatial data and location-based services more accessible via non-visual interfaces. Benjamin is one of the lead developers of the PocketNavigator, one of the selected HaptiMap demonstrators. This Android-based application is able to guide a user to an arbitrary destination by providing tactile feedback.

The evaluation of these mobile navigation aids is a challenge. Thus, Benjamin's research focuses on mobile 'in the wild' context sensing and evaluation techniques, especially in the domain of mobile navigation, exploration and location-based applications. Until now he focused on the quantitative recognition of certain navigation behaviours. His future research plans focus on the contextual enrichment of qualitative user observation techniques. Benjamin has more than 10 peer-reviewed workshop and conference publications. Further, he organised the international OMUE (Observing the Mobile User Experience) workshop and did reviews for several HCI conferences, like CHI, MobileHCI, UIST, INTERACT, and Ubicomp. He was selected for a Microsoft travel grant to participate and present his work at the SenseCam 2010 symposium.