Using App Stores to Conduct Longitudinal Research

Martin Pielot

OFFIS Institute for Information
Technology, Germany
pielot@offis.de

Wilko Heuten

OFFIS Institute for Information
Technology, Germany
heuten@offis.de

Susanne Boll

University of Oldenburg, Germany
Susanne.boll@informatik.uni-
oldenburg.de

Abstract

Conducting longitudinal user studies can be highly
resource consuming. Thus, research groups have
investigated to use emerging application distribution
channels, such as Apple's App Store or Google's
Android Market to bring the research to the user. By
turning an app into an experimental apparatus we can
study usage patterns that occur in-situ and for a long
period of time. However, these kinds of studies are still
hardly understood and come along with a lot of
challenges that threaten such studies' internal validity.
In this paper, we report from two case studies from our
research group and elaborate the challenges we
encountered.

Keywords

Longitudinal Research, App Stores, Handheld
Computing

Background and Motivation

Conducting user studies is a corner stone of increasing
our knowledge about how people interact with
technology. When studying new interface artifacts this
often happens in lab or field studies. Participants are
invited for an hour or two and use given interface
artifacts to fulfill given tasks.

This common type of study set-up has the disadvantage that the study's context and the tasks to fulfill are specified by the experimenter. The participants interact with the studied interface artifacts because they are asked to, but not out of a necessity. Further, the participants do not have the time to become experts on the interfaces and develop their own usage strategies. Thus, the results of such short-term studies in the lab or the field may lack *external validity*, i.e. they cannot necessarily be applied to real usage of the studied interface artifacts.

One way to solve these issues is to release interface artifacts "into the wild" as part of a longitudinal study and have them used by people as part of their daily life. This allows participants to become familiar with the artifact and develop their own usage strategies. Further, users can use the artifact in-situ, which we consider as usage when there is a true need to use it. This allows studying different kinds if usage contexts, in which the artifact will really be used.

However, organizing such a study in a traditional way, where each participant is supervised individually, can be highly time and resource consuming. Each participant has to be briefed, trained, and debriefed. Sometimes, collecting data can require additional resources, e.g. when they have to be obtained from direct observation.

Research Using Apps and App Stores
As a solution to this, several research groups [1,2,3,4,5], including ours, have investigated using emerging application distribution channels, such as the Apple's App Store, to study interaction techniques and

interface artifacts with hundreds to hundreds of thousands of people.

Hit It! For example, Henze et at. [1] developed a game for Android phones where people have to pop bubbles by hitting them on the touch screen. Henze et al. collected more than 100,000,000 touch events and found systematic biases. They proposed a correction function and by integrating it into an update of the game they could show that the function significantly increases the users' touch accuracy.

PocketNavigator: In our recent work [5], we developed a pedestrian navigation application called PocketNavigator and released it via Google's Android market. Our aim was to study the effect of vibro-tactile feedback on the traveler's level of distraction. By inferring the level of distraction from the device status and sensors we could show that users are significantly more likely to be less distracted when the vibro-tactile feedback is enabled.
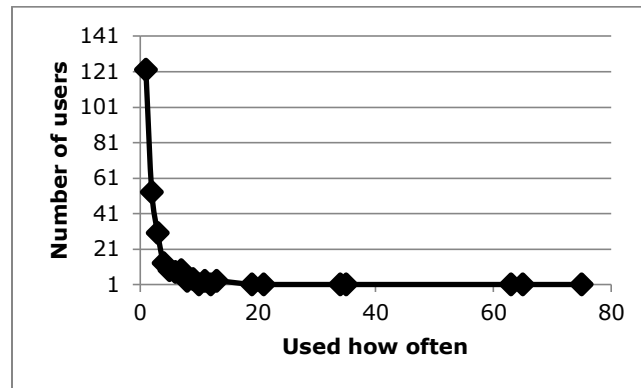
Challenges
Based on above two studies and further previous work (see [2] for a good overview) we found that such studies can successfully be used to investigate interface artifacts in-situ and with a large number of users. However, we also found a number of issues that make these kinds of studies challenging.

Many short-time users
Regarding long-term usage, we found that the majority of the users run the application for a short time only. In the cases of the game and the navigation application described above, many users just played a single level

Figure 1 Screenshot of the PocketNavigator - a navigation application that we used to study vibro-tactile feedback "in the wild".

of a game or just started the navigation application once, presumably for testing.

Figure 2: Number of trips per user with the PocketNavigator - most of the users just did a single trip. Only few used it more than ten times.

This becomes a challenge if the study intends to investigate long-term use and experienced users, not first-time use and beginners. Thus, it is mandatory to meet the expectations of the users. If the quality of the actual application is poor users may just delete the application. If it meets the basic expectations it is still necessary to motivate long-term use. In any case it is necessary to log the amount of usage and filter out short-time users, if long-term factors are investigated.

Data Collection
Another challenge we regularly face is how to collect the necessary data for the study. The only existing feedback channels for users are leaving comments in the Android Market or send emails to the developers. From our experience this feedback is largely useless. The comments from the market are typically short and

unspecific and oftentimes are just complaints[1]. Mails are mostly used to request new features. McMillan et al. [4] tested two promising methods to collect more meaningful qualitative feedback. They offered incentives, such as badges or achievements, for giving qualitative feedback and tried to contact users directly.

In our work we investigated using the mobile device's sensors to gain an understanding about relevant aspects. In the case of Hit It!, Henze et al. [1] collected the x and y coordinates of each touch event together with the location and diameter of touchable items and other relevant aspects of the game status. In the case of the PocketNavigator we went even further and e.g. used the device's magnetometer and accelerometer readings to infer how people use the device (e.g. if it is likely that they are looking at the display and thus are distracted from the environment). Still, we face the challenge that these finding contain a lot of uncertainty.

Unpredictable usage
One particular source of uncertainty is the unpredictable use of the applications. For example, we found many instances where user started the application and then just left the device lying on the flat surface for some time. Thus, it has to be expected that results always contain artifacts or noise, which makes it challenging to obtain valid results from the data.

One solution is to carefully investigate the data and apply filters where appropriate. In the case of the PocketNavigator, we e.g. filtered out all data where the user was not moving according to the GPS signal. Further, applications should be designed to make

---

[1] http://pielot.org/2011/03/24/android-user-hate-parade/

unforeseen usage less likely. For example, the PocketNavigator offers a map and therefore does not necessarily have to be used as a navigation aid. One solution might be to remove the map and only provide navigation instructions, however, this may collide with the first challenge of meeting the users' expectations.

Conclusions

Publishing applications via app stores, such as Apple's App Store or Google's Android Market is slowly being recognized as a potential option to study interface artifacts in the wild. To study long-term usage, however, there are a number of challenges that have to be addressed in order to obtain valid results. This includes motivating long-term use, collecting valid data, and ensuring that only the desired usage is studied. We believe that addressing these challenges will create great opportunities to conduct long-term studies in the large while keeping the resources at a reasonable level.

Bio

Martin Pielot is a research assistant and PhD candidate at the OFFIS Institute for IT in Oldenburg, Germany. His PhD focuses on conveying spatial information by vibro-tactile displays to address situation-induced impairments when using handheld devices on the move. One of the application areas focuses on providing navigation instructions via

vibration patterns. To study the effects of these tactile cues in-situ with a large number of participants he investigates using the Android Market as a tool to conduct large-scale studies.

References
[1] Henze, N., Pielot, M., Poppinga, B., Schinke, T. and Boll, S. My App is an Experiment: Experience from User Studies in Mobile App. *International Journal of Mobile Human Computer Interaction* 3, 4 (2011), 71-91

[2] Henze, N., Rukzio, E. and Boll, S. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proc. MobileHCI '11*, ACM Press (2011)

[3] Cramer, H., Rost, M., Belloni, N., Chincholle, D., Bentley, F. Research in the large: Using App Stores, Markets and other wide distribution channels. In *UbiComp research, Extended Abstract & workshop* organised at UbiComp 2010.

[4] McMillan, D., Morrison, A., Brown, O., Hall, M. and Chalmers, M. Further into the Wild: Running Worldwide Trials of Mobile Systems. In *Proc. Pervasive 2010*.

[5] Pielot, M., Poppinga, B., Heuten, W. and Boll, S. PocketNavigator: Studying Tactile Navigation Systems In-Situ. In *Proc. CHI 2012*, ACM Press (2012).