

Experiences of Mass Participation in Ubicomp Research

Alistair Morrison, Stuart Reeves, Donald McMillan, Matthew Chalmers

Department of Computing Science, University of Glasgow

{morrisaj, stuartr, donny, matthew}@dcs.gla.ac.uk

ABSTRACT

The emergence of ‘app stores’ as a means of distributing software applications on a number of mobile platforms is a relatively recent phenomenon. Ubiquitous computing (ubicomp) researchers are only just beginning to run trials that take advantage of the large numbers of users these distribution models can offer. The relatively easy access to such a potentially wide audience could radically alter the nature of many ubicomp trials, yet as the practice is so new, the field has not yet developed a set of guiding principles or an understanding of good practice. In this paper we share our experience of running trials of several iPhone applications deployed in such a manner, describing our findings and offering recommendations to other researchers planning to use app store-style distribution.

Author Keywords App stores, distribution, mass participation trials.

ACM Classification Keywords H.5.2 User Interfaces: Evaluation/methodology

General Terms Experimentation, Human Factors

INTRODUCTION

The use of online software repositories (‘app stores’) in ubicomp research is a relatively recent phenomenon. Although the Apple App Store launched in July 2008 and in its first two years has had over 5 billion downloads [1], ubicomp projects are only beginning to utilise such a repository as a means of deploying research applications [2]. As such, the ubicomp community is yet to come to a consensus on a set of guidelines for conducting such trials, or agree on firm ideas of what constitutes good practice.

In our research into mass participation ubicomp trials, we have released a number of research applications using the app store model of distribution, experimenting with different types of application and different methods of conducting a trial. In this paper we describe some of our experiences, exploring issues such as the number of users that researchers can hope to attract, demographics of users and the ease with which feedback can be elicited from the user base. We offer our conclusions on what has been successful in our trials, and what has been less so, and we suggest recommendations for others working in this area.

SUMMARY OF APPLICATIONS

Several applications are discussed in this paper and will be briefly introduced here. Hungry Yoshi [2], a game that uses

wifi infrastructure as a game resource, was the first application we released through the app-store method and it currently has around 40,000 registered users, many of whom have engaged in the trial process through completing feedback questionnaires. Although it has had a large number of downloads, not everyone went on to play the game, as it requires users to find ‘fertile’ areas to play with a good number of wireless base stations. Packer is a simple memory game designed to be playable anywhere without such worries of surrounding wifi infrastructure. More recently we’ve released WorldPaint, a game to draw patterns on a map based on participants’ movements, which experimented with allowing audio as well as textual feedback from users. Our most recent application is World Cup Predictor, a game designed to run alongside the FIFA World Cup, which had £500 of prizes for top players and tried to encourage social interaction with other players through Bluetooth-based data transfers. All apps were free to download via an APT-based repository [3].

These systems were all instrumented with the SGLog framework [4] which regularly uploads usage logs to our servers, providing the basis for the statistics and analysis in the following sections.

USER NUMBERS

An obvious question while planning a trial using an app store distribution mechanism is the number of users researchers can expect to download and use their trial software.

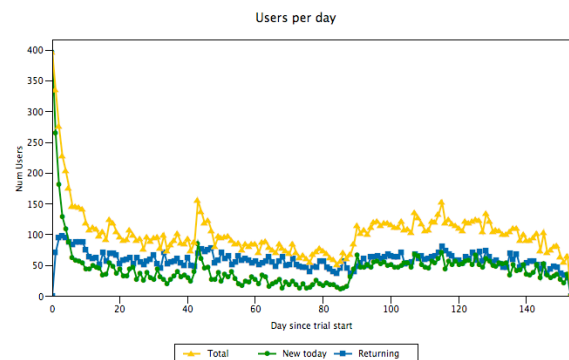


Figure 1. Number of users running an application by day.

Figure 1 shows a graph of the number of users running the Packer application per day since release. Separate trends show number of new users that day (green), users who have run the application on an earlier day (blue) and the sum of these two values (yellow). It can be seen that the highest number of users is seen on the day of release (400), gradually declining to a plateau of around 100 per day.

Increases up to around 150 occur on the days an update to the software was released. There are usually roughly similar numbers of new and returning users each day.

Although the specific numbers might vary, this general pattern is something we've regularly observed across almost all applications. The initial peak is most likely due to the application being at the top of the 'newest applications' list and therefore more visible to people browsing the store. We suggest that the increase in users on new version releases are because an update to an existing application also bumps it back to the top of the 'new' list, and because people owning the application will perhaps run it again to see what has been updated.

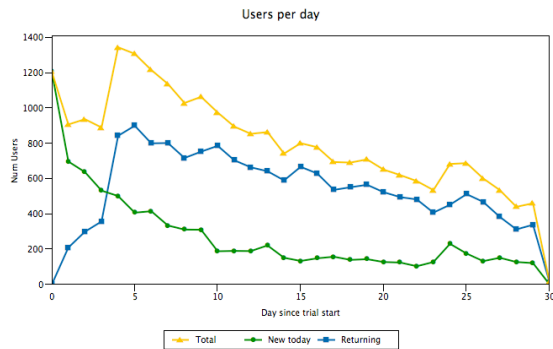


Figure 2. Number of returning users is higher for an application running alongside an ongoing event (the FIFA World Cup).

The one exception to this pattern is the World Cup Predictor application, as shown in Figure 2. Although it might be expected that an application based exclusively around a single sport might have less universal appeal than for example a puzzle game, we found that we had much higher user numbers for this application compared to the others, especially the number of returning users. The two factors that distinguish this application from the others are its tie to an ongoing real-world event that users might already be interested in, and the fact that we offered a cash prize to the winner. Feedback from an in-app questionnaire indicated that 64% of users felt that football was very important to them, with only 7% stating that they had little or no interest in the World Cup, suggesting that users were generally being attracted due to an interest in the event rather than solely chasing an opportunity to make money.

WHAT IS 'A USER'?

In research papers it is common practice to state the number of users involved in a trial. Yet an interesting question emerging from this sort of trial is what exactly constitutes a user of an application. In traditional trials, the definition of a participant is usually more clear, with applicants perhaps responding to a recruitment announcement, being supplied with a device to use and being paid for their time. In trials that utilise an app store-style distribution method, this becomes more complicated.

The charts presented in Figures 1 and 2 count the number of people who have launched the application, but this does not fully explain how intensively a system is being used. Is someone who runs a system once for 20 seconds before

uninstalling it 'a participant' in the trial? If not, how does a researcher choose a threshold for activity before the user can be said to be engaging in software use; what makes an 'active user'?

As well as use of the software, it might also be important to consider a participant's engagement with the trial itself. Many of our applications have included feedback mechanisms through which researchers can target specific questions to users, as explained below. Trial engagement could also come in the form of telephone interviews, emailed questionnaires or communication through social networking sites.

As an example of the different ways users can be counted, some numbers are provided from our trial of Hungry Yoshi [2]. Stats provided from the online repository state that the application has had 182,714 downloads. This figure includes software updates, so the same user might be included up to 8 times if he or she has downloaded every new version of the application. To count unique users, our SGLog database has recorded that 98,556 people have launched the application, although this number might be lower than the true value if the application was run while network connectivity was unavailable on the user's device, as SGLog would not have been able to upload to the server. On first launch, the user is presented with terms and conditions and asked to sign up for an account in order to play the game. 36,169 completed this registration process. Considering only the users who managed to score any points in the game (a non-trivial task that may involve physically walking to an area to collect 'fruit' then taking it to a Yoshi waiting at a different real-world location) the figure drops to 4,134, and looking at players who played on 5 or more different days the number is down to 3,080. It can be seen therefore that simply reporting the number of downloads an application has had is not a particularly informative statistic on its own.

ENCOURAGING USE

Having discussed user numbers, this section discusses different methods we have experimented with in encouraging sustained use of trial applications.

In more traditional ubicomp trials, participants are often being compensated financially for their time. There could be a set number of tasks a participant is expected to complete to qualify for payment, or the pressure to use an application might be more implicit - that the user is part of the trial and they might feel that they have a duty to use it. With an application downloaded from an app store, researchers are probably not going to pay for participation and a user might feel less obligated to put in 'token' hours of use or otherwise engage with the trial process.

Depending on the research questions under investigation, researchers might therefore have to think of other means to motivate use. Most of the applications we have trialled have been games and, as such, it is expected that fun will be a motivating factor. It is also often possible to introduce a competitive element. Several of the games we have released have included a global scoreboard, where players

can compete to rise in the rankings. Our interviews with users of Hungry Yoshi indicated that scoreboard position was an important factor for players, with one user for example stating *“I definitely don't want to go anymore back. Actually in a week I think I will go one more up”*, showing that he was defensive of his current position and was actually making reasonably longterm plans about how he would rise up the table.

Although it would not be feasible to offer to pay every user who downloads and runs a trial application, researchers can still offer financial rewards as prizes for success or for fulfilling certain obligations. For example, in the Hungry Yoshi trial, participants were paid if they agreed to be interviewed over the telephone about their use of the game. A feedback question was put into the game to gather email addresses of those willing to be interviewed in this way.

SOCIAL APPLICATIONS

Although the app store distribution mechanism provides a simple means to deploy an application to a large user base, the situation becomes more complicated if the software under examination has a social aspect. Whereas in a traditional deployment groups of friends or colleagues might be recruited en masse for such a trial, with app store style distributions it cannot be guaranteed that groups of users will all download the application.

Two of our trialled applications have attempted to provide social elements in different ways. In Hungry Yoshi, users were provided with an option to sign into the game using a Facebook account, in order to share progress with friends or to chat with friends or developers about the game. 16,735 of 36,169 registered users chose this option, which is 46% of the user base.

Using a different approach, the World Cup Predictor game offered additional bonus points to players for playing a ‘head-to-head’ game with a friend where predictions were swapped locally over a Bluetooth connection. Despite the strategic advantage that could be gained from performing these head-to-heads, only 45 out of 3,847 users who played the game did so, which is only slightly more than 1%. This low number was in spite of the fact that there was a prize for winning the league and the fact that those players engaging in head-to-head activity were gaining an advantage - 2 of the overall top 3 players were among the 45 Bluetooth users.

We have also explored how application usage might spread through social groups, asking players of Hungry Yoshi whether they had told their friends about the game or encouraged others to play. Roughly a third of responses indicated positively that players had spoken about the game with their friends, for example *“Ive recommend the game to a lot of my peers. They love it go yoshi!”*.

These results seem to indicate that users are keen to add social aspects to their applications, but are more likely to engage in activities such as Facebook, with its flexibility to be used anywhere and at any time, than those that require users to be co-located.

DEMOGRAPHICS

Several of our applications have included short questionnaires asking demographic information of the users such as their age and gender. Answers could always be provided by pressing one button from a small selection or moving a slider. These questions were presented to the users when asking for a username to use on the scoreboard, yet were not mandatory and users were not rewarded in any sense for answering. Nevertheless, a high percentage of the participants answered these questions, providing a reasonable insight into the characteristics of the user base. To illustrate, Figure 2 shows the spread of the ages of the World Cup Predictor application, where 80% of the users provided an answer to the question.

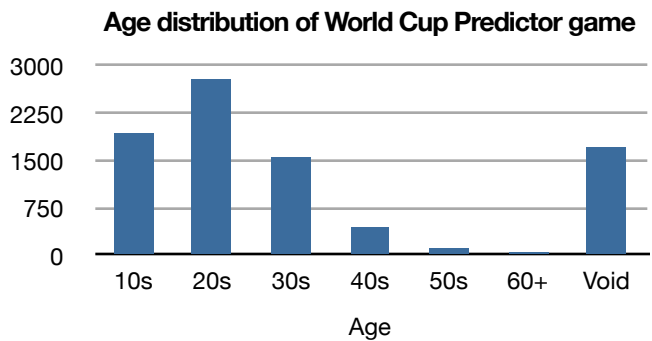


Figure 2. Distribution of ages of users of the World Cup Predictor app. 20% of users failed to provide an answer.

It has been typical of the trials we have run that most users are aged between 15 and 35. The gender of users we see running our software tends to be biased towards males, as shown in Figure 3. Perhaps unsurprisingly this trend is most apparent in our football prediction application.

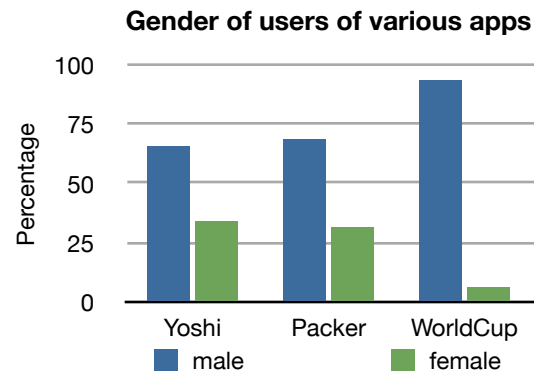


Figure 3. Gender distribution percentages for several apps.

As always, care must be taken in interpreting these results in that there might be a bias inherent in those people who are choosing to answer.

GATHERING MORE DETAILED FEEDBACK

In addition to simple demographic information, we have experimented with asking more complex questions and tried to elicit more detailed responses from users, allowing them to enter free text into a form on the application using the iPhone’s soft keyboard.

Although not answered by the same high percentages of users as the more simple demographic questions, these forms of information gathering were still used by many participants and have proved very useful. In Hungry Yoshi, players were rewarded with in-game ‘tokens’ for answering questions, and 6,115 players did so, which is 17% of registered users and actually more than the number of players who ever scored a point in the game. In the World Cup Predictor, where a feedback section was included but no rewards offered for answering questions, 932 or 11% of registered users gave at least one response.

We have had great benefit from information gained through these channels, being made aware of numerous bugs, received and subsequently implemented feature requests and setup interviews with participants. Users were also happy that their opinion was being valued, with one stating “*I find it really nice that [you are] contacting me and asking me my opinion. I guess it’s a really nice thing.*”

Most app stores on various platforms also allow users to enter reviews, which will be publicly shown alongside the application in the download area. Users seem to use this mechanism for two purposes - as advice for other people on whether the application is good, but also in many cases as a feedback channel to the developer. Yet despite this standardised mechanism for providing comment, users appear far more likely to use the in-app feedback mechanisms than these store-provided ones. As an example, the World Cup Predictor has had 2 reviews in the store compared to, as mentioned, comments from 932 users in the application, despite there being no reward for doing so. And it is interesting that many of the comments received read like a review one might see on a store, for example “*Everything is good about this app. Very useful!*”

Our WorldPaint application also provided participants with the opportunity to leave feedback via audio comments which were recorded on the device then uploaded. We found that users were less likely to choose this option, leaving 397 audio comments as opposed to 1746 text comments. One reason for this disparity could be the far longer upload times for audio files.

Language issues likely play a part in not seeing greater uptake of the feedback features, as many of the responses received were in foreign languages.

CONSENT AND PRIVACY ISSUES

In running our trials, every application displays a consent form on first launch displaying terms and conditions which a user must agree to before he or she can proceed further and before any log data is captured. These terms contain descriptions of all the forms of data that will be logged, as well as information on how the data will be stored and accessed. As these are worldwide trials, it cannot be expected that every downloader will be fluent in English, so we ensure that this form is presented in at least four major languages. It is also made clear to users at this point that any data collected on them will be destroyed on request, with an email address provided to facilitate this service. To date no such request has been received.

RECOMMENDATIONS FOR TRIALS USING LARGE SCALE DISTRIBUTIONS

To conclude, we summarise a series of recommendations based on our experience of running several ubicomp trials using app store-style distribution.

Engage with multiple interpretations of the active user. Static or singular metrics for what is considered an ‘active’ user were problematic for us. Instead, we found that we had to constantly redefine a multiplicity of definitions for what was an ‘active’ user, resulting in our interest in increasingly sophisticated categorisation schemes.

Consider the ways in which different levels of user activity may be accounted for in the application design. For instance, can your application design make a virtue of a large number of ‘single use’ participants?

Use the distribution mechanism to shape user flow. We found that application store features, such as a ‘newest’ section that includes updated apps, offered the increased potential to catch the attention of new users. App owners who have stopped using it may also be prompted to launch it again to see what’s changed. In this way we could optionally impact the flow of users to our application.

Integrate data collection and evaluation mechanisms into the context of the application. The inclusion of a feedback section in an application, where users can answer questionnaires or leave open comments is a very useful feature. We saw a higher uptake of this feature when points were offered for responses, and users seemed happier to enter text than audio comments. The response rate was much higher still when simple multiple choice buttons could be pushed to answer questions.

The tendency for individualist use rather than co-located social use of applications may be high. We found little uptake of Bluetooth head-to-head features of the World Cup app, even when large point bonuses (and therefore higher chance of prizes) were offered to entice use. Social functionality through Facebook features was far more widely used. A crucial question to be answered in future work is the relationship between extensive social use and a critical mass of users, e.g., is more social use of an application strongly related to the ownership density?

REFERENCES

1. CNNMoney, <http://tech.fortune.cnn.com/2010/06/07>
2. McMillan, D., Morrison, A., Brown, O., Hall, M., Chalmers M. Further into the Wild: Running Worldwide Trials of Mobile Systems. *Proc. Pervasive* (2010), 210-17.
3. McMillan, D. iPhone Software Distribution for Mass Participation Trials. *Research in the Large*. CHI Workshop (2010).
4. Hall, M., Bell, M., Morrison, A., Reeves, S., Sherwood S., Chalmers, M. Adapting UbiComp Software and its Evaluation. *Proc. ACM Engineering Interactive Computing Systems* (2009), 43-148.