

It's a Jungle Out There: Practical Considerations for Evaluation in the City

Melanie Kellar, Derek Reilly, Kirstie Hawkey, Malcolm Rodgers, Bonnie MacKay,
David Dearman, Vicki Ha, W. Joseph MacInnes*, Michael Nunes, Karen Parker,
Tara Whalen, Kori M. Inkpen

Faculty of Computer Science
Dalhousie University
Halifax, NS Canada

* Department of Math and Computing Science
Saint Mary's University
Halifax, NS Canada

inkpen@cs.dal.ca

ABSTRACT

An essential aspect of mobile and ubiquitous computing research is evaluation within the expected usage context, including environment. When that environment is an urban center, it can be dynamic, expansive, and unpredictable. Methodologies that focus on genuine use in the environment can uncover valuable insights, although they may also limit measurement and control. In this paper, we present our experiences applying traditional experimental techniques for field research in two separate projects set in urban environments. We argue that although traditional methods may be difficult to apply in cities, the challenges are surmountable, and this kind of field research can be a crucial component of evaluation.

Authors Keywords

Mobile computing, ubiquitous computing, field research, methodology, evaluation.

ACM Classification

H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces: Evaluation/methodology

INTRODUCTION

Observation in context is a fundamental part of mobile and ubiquitous computing research practice. This is because the systems and techniques developed are intended to support day-to-day activities in a specific environment (or range of environments). Field research strategies allow researchers to explore environmental and contextual impacts, but make it difficult to achieve the precision and control of laboratory experiments [8, 9]. We often accept these limitations in order to observe behaviour in a natural setting.

When the context of observation includes manoeuvring in an urban environment, unique methodological challenges arise. Social norms and activity patterns in the city are distinct from those of office or home environments, and can exert a strong influence on the behaviour of participants.

Urban environments, such as parks and shopping districts, are shared, multi-purpose spaces, difficult to predict and impossible to control. Mobility makes strong demands of observers. The fidelity of observation and measurement available in this context is severely reduced relative to controlled labs or even office environments.

Evaluation in city spaces can certainly afford realism. However, this does not come automatically: seemingly innocuous choices in a study's design can dramatically impact the "reality" of the situation under study. Also, while there is an overall trade-off between measurement, control and realism in urban settings, some traditional measurement strategies can be adapted more readily than others. It is important to understand at a detailed, practical level why a strategy works or fails in such environments.

In response to these methodological challenges, research strategies have been proposed that sacrifice or severely curtail direct observation [3, 5, 6]. However, it is possible to conduct useful experiments in the city. In this paper we describe our experiences conducting such research in two separate projects. We illustrate challenges and benefits of taking techniques out of the lab and "into the wild" of the city streets.

BACKGROUND

According to the classification scheme provided by McGrath [9], evaluation methodologies in natural settings include case studies, field studies, field experiments, and experimental simulations. There are advantages and disadvantages to each of these approaches and previous research has highlighted use of these techniques from the perspective of mobile, ubiquitous computing.

Despite general agreement on the importance of field research for mobile computing, over 70% of related HCI papers published between 1998 and 2002 considered in [7] involved only a lab experiment. Lab experiments are often beneficial when evaluating small pieces of a complex problem (or system) [3], but should be combined with field research to better understand the impact these systems have in a normative environment. Although field studies or

experiments are attractive options, data collection and analysis is often problematic. This was very apparent in Moran et al.'s [10] work where the researchers experienced severe difficulty in analyzing audio data.

Pragmatic adaptations of field research methodology have been developed in response to difficulties (experienced or perceived) in conducting research in the field. For example, Experience Clip [6] is a technique to overcome the intrusive influence of a researcher in a mobile context by encouraging users themselves to take short video clips. Another technique [4] places a researcher as an actor with a participant in their daily lives, enabling understanding of context and aiding in the evolution of new designs based on experience. Finally, researchers such as Intille et al. [5] have explored ways to facilitate data collection using context-aware experience sampling techniques. Each of these approaches is valuable, but they are not intended to replace other field research methodology.

TWO URBAN NAVIGATION EXPERIMENTS

We briefly describe two field studies we conducted, which will be used to illustrate challenges of urban field studies.

Rendezvousing

Our rendezvous study [2] investigated the impact that location-aware handheld technology would have on the act of rendezvousing. Situated in a busy, downtown shopping district, 24 pairs of participants completed three rendezvous scenarios using either cell phones, handheld computers, or both devices. The map application on the handheld showed their current location, that of their partner, and the rendezvous location. We mimicked a wireless location-aware system using a Wizard of Oz approach involving a researcher (wizard) assigned to each participant. The wizards relayed locations to each other using 2-way radios, and updated locations on participants' maps using a separate, Bluetooth-enabled handheld. One observer also followed each participant, recording field notes and giving instructions during the study (figure 1a).



Figure 1. An observer and “wizard” follow a participant in the rendezvous study (a). An observer (on the left) prepares to pull a video camera out of his backpack during the shared annotations study (b).

Shared Annotations

In this study, we explored the benefits of sharing annotations across mobile devices for co-located users. Three pairs of participants used handheld computers during

the City Chase¹, an organized event in which teams raced against each other to navigate a city, solving clues in order to find race “pit-stops” that presented challenges to complete. Our three pairs were equipped with electronic maps and bus schedules on their handhelds, along with paper equivalents, to help them in the race (figure 1b).

EXPERIMENTS IN THE CITY

Experimental Design

The rendezvous study was an *experimental simulation*, designed to observe behaviour in the context of rendezvous scenarios, as defined by Colbert [1]. However, the unpredictable city environment made it difficult to simulate each scenario precisely as planned, (e.g., person A reaches rendezvous location before person B). Although this ‘unpredictability’ was challenging, it did enable observation of a broad range of rendezvousing behaviour.

In the shared annotations study, we evaluated our prototype in the context of a real event (the City Chase). This gave us a genuine activity to explore the technique. However, we were given only a high-level description of the race prior to the event, so we prepared by envisioning potential race scenarios. The study was designed as a *field experiment*. In order to maintain the reality of the event, we outfitted each team with paper maps, cell phones, and other potentially useful tools. Two teams were given the shared annotations software, while a third team was given handhelds with the same information, but without shared annotations.

Piloting and Feasibility Testing

Coding sheets were created for the rendezvous study on the assumption that entering structured observations would be easier for an observer in motion. Pilot testing quickly showed that detailed observations were extremely difficult to capture on the street. Coding sheets were reformulated to capture high level categorizations like “patterns of map use”. To reduce the amount of paper to manage, coding sheets were integrated into scripts and checklists.

We conducted a mock scavenger hunt prior to the City Chase event. Because the details of the actual event were unknown, we focused our efforts on determining the general structure of the experiment. As a result of the trial, we decided to assign an observer to one of the shared annotations pairs, leaving the remaining two pairs subject to (less intrusive) audio capture and software logging only.

Audio was used in both the rendezvousing and shared annotations studies to record verbal interactions between the participants. Mock scenarios were conducted to measure audio quality and place recording equipment on the person.

Running the Experiments

When we ran the experiments, we were confronted by a range of issues affecting experimental control and our

¹ www.thecitychase.com

ability to observe behaviour. While some of these issues may seem obvious, especially in hindsight, they highlight the external factors that impact both research and adoption of mobile, ubiquitous computing technologies.

Software

Connectivity issues with the Bluetooth devices in the rendezvous study meant researchers had to interrupt sessions to reset the equipment. In two cases, the interference was great enough that participants commented that it affected their behaviour and these sessions had to be discarded. Software failure prevented one participant pair from using the handhelds at one point in the shared annotations study, which in turn discouraged further use.

Materials

The lack of a “home base” had a considerable impact on the rendezvousing study. Equipment and paperwork had to be carried throughout the experiment. There were no power outlets, which meant that battery power had to be carefully managed during the long study days. Experimental conditions were assigned, in part, based on the level of battery power left in the devices. We used benches for interviews, clipboards for paperwork, and learned the value of pens that did not have to remain vertical.

In the shared annotation study, participants were fitted with microphones and an extra waist pack to carry their handhelds. Although the extra gear was somewhat cumbersome during certain competitive “race points”, it did enable collection of audio data without the participants having to worry about the technology.

Social considerations

The general public influenced participant behaviour and the flow of each study. Participants had to dodge pedestrians—sometimes unsuccessfully—as they navigated using their maps or handhelds. Curious passers-by sometimes stopped to watch, or to ask researchers or participants what they were doing. The City Chase became an increasingly social event as the race progressed, and participants were completely caught up in the atmosphere. Being in a public space increased the potential for embarrassment and feelings of self-consciousness on the part of the participants, especially in the rendezvous study, where participants were followed by an entourage of observer and wizard. City encounters also occurred with people on the street such as buskers, artists and panhandlers.

Weather

During the rendezvousing study, participants were rescheduled when rain was an issue. However, wind and sun complicated the study. Wind made it difficult for participants and researchers to handle paper forms. Bright sunlight made it difficult at times to view the handheld displays. Researchers and participants also needed to be careful of exposure and dehydration in the heat of the sun.

Other environmental factors also created problems: in an urban park, tree sap dripped on the equipment.

Rain was a bigger problem in the shared annotation study, as it rained for the entire morning of the race. Because we were participating in a real event, we could not reschedule. The handhelds went into baggies, the cell phones and microphones got wet, the observer’s video camera was wrapped in plastic, and paper materials disintegrated.

Audio and video

In both studies, it was difficult to capture quality audio recordings due to background noise, which was in general far worse than that encountered during feasibility testing and pilots. While participants walked on sidewalks, their recorders picked up third party conversations. Clip-on microphones captured jostling and fabric rustling as we ran about during the City Chase. Environmental noises such as construction, tour bus commentaries, large trucks, and traffic were continuous. In the case of the rendezvous study, city noises often drowned out the voice recordings as well as the researchers as they gave instructions to the participants. Recording video in the fast-paced environment of the City Chase meant that video data was shaky and viewing angles were less than ideal. Although the video quality was poor, it was still a critical component since we were attempting to examine aspects of collaboration among users of mobile devices, and capture of fine-grained details such as glances and gestures was important.

Mobility

In addition to related audio and video concerns, high mobility and time pressure influenced both technology use and observation. In the rendezvous study it was difficult to stay close to participants in crowded areas. We frequently needed to remind participants to walk at a moderate pace and obey traffic signals so that the observers could safely keep up. In one instance, a participant became completely separated from the researchers after darting out as a traffic light was changing.

The unconstrained mobility also made it difficult to monitor interactions with materials and to interpret gaze: was the person looking at the handheld or down the street? Field notes taken while walking were terse and messy, and difficult to transcribe. Observers generally took notes when the participants were stationary, which was not necessarily when something was observed. In the City Chase event, direct observations were even more challenging since the participants were often running. Regardless, handhelds were used by participants only while riding the bus or ferry.

Collecting and Analysing Results

Most of the data collected during the rendezvous study was useful during analysis; however, it was often not used as originally intended. Software logging confirmed where people set rendezvous spots, and was cross-referenced to decipher observer notes and shed light on comments made in interviews. Timing data was recorded, but since the

scenarios were difficult to control, timings were highly variable. Timing data was instead used to provide “landmarks” when describing specific rendezvous trials. Audio recordings of participants were at times beyond reconstruction. Background noise drowned out participant comments making the audio impossible to transcribe. In some instances transcription was possible as one participant’s comments could be discerned from their partner’s recordings. Researcher observations proved to be useful, especially in combination with participant evaluations and interviews. Observers supplemented participant comments that were incomplete or misleading.

The shared annotations study was marked by an almost complete absence of prototype use across all pairs. As such, most video and audio data was discarded, and software logs were non-existent. It initially appeared as though the demands of the real event overtook the needs of the experiment. However, participants acquired a real, visceral sense that the technique was ill-suited to the City Chase context. Post-hoc debriefings and careful reflection was therefore critical to our interpretation of the result. Through this we gathered ideas about what would have been more useful, and how the prototype could be changed. It gave us a clearer sense of the parameters of design for further study, and of the challenges in using handheld devices in “high octane” environments.

CONCLUSION

It has been suggested [3] that experimentation in the field represents a good balance between field studies and lab experiments. While this style of evaluation can be complex to run and analyze, the obstacles that make these methods challenging are also those aspects that are crucial to understand, such as a dynamic, noisy environment, mobility, and context-dependant behaviour. Direct observation and experimentation, however problematic, are effective tools that contribute to this understanding.

The experimental simulation style of the rendezvous study allowed researchers to observe behaviour that may not have been apparent in a more controlled setting. In any research, there is always the possibility that participants try to anticipate and fulfill researcher expectations. However, once participants are immersed in a natural environment, their natural instincts and behaviours may emerge, as illustrated by our participant who ran to make the light. In this study, the urban environment significantly reduced experimental control in many ways and limited detailed observation, but lent a realism that was felt by the participants and allowed evaluation of the technique in its intended context.

While the shared annotations experiment failed to produce empirical results, it allowed us to evaluate factors that may have contributed to its non-use, both from a methodological perspective (was the chosen event an appropriate context of use) and from a technology adoption and use perspective (is there still too much overhead to make handhelds feasible to

use in mobile environments). So while the field experiment failed *per se*, the lessons learned are arguably more valuable than those attainable in an artificial lab setting, and can be used to direct further study of this technique.

We have reflected upon our experiences applying standard field research techniques when evaluating technology in an urban context. Dynamic and unpredictable, urban environments seriously challenge experimental observation and control. Yet, as our experiences demonstrate, there are also tremendous insights to be gained.

ACKNOWLEDGEMENTS

Thanks to the members of the EDGE Lab for their support. This research was funded in part by NSERC.

REFERENCES

1. Colbert, M. (2001). A Diary Study of Rendezvousing: Implications for Position-Aware Computing and Communications for the General Public. In *Proc. of Group 2001*, Boulder, CO.
2. Dearman, D., Hawkey, K., and Inkpen, K. M. (2005). Effect of Location-Awareness on Rendezvousing Behaviour. In *Extended Abstracts of CHI 2005*, Portland, OR. (in press)
3. Goodman, J., Brewster, S., and Gray, P. (2004). Using Field Experiments to Evaluate Mobile Guides. In *Proc. of 3rd Annual Workshop on HCI in Mobile Guides*, Glasgow, Scotland.
4. Iacucci, G. and Kuutti, K. (2002). Everyday Life as a Stage in Creating and Performing Scenarios for Wireless Devices. *Personal and Ubiquitous Computing*, 6(4): 299-306.
5. Intille, S. S., Tapia, E. M., Rondoni, J., Beaudin, J., Kukla, C., Agarwal, S., et al. (2003). For Studying Behavior and Technology in Natural Settings. In *Proc. of UbiComp 2003*, Seattle, WA. 157-174.
6. Isomursu, M., K., K., and Väinämö, S. (2004). Experience Clip: Method for User Participation and Evaluation of Mobile Concepts. In *Proc. of Participatory Design 2004*, Toronto, Canada. 83-92.
7. Kjeldskov, J. and Graham, K. (2003). A Review of Mobile HCI Research Methods. In *Proc. of Mobile HCI 2003*, Udine, Italy. 317-335.
8. Kjeldskov, J., Skov, M. B., Als, B. S., and Hoegh, R. T. (2004). Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In *Proc. of Mobile HCI 2004*, Glasgow, Scotland.
9. McGrath, J. E. (1995). Methodology Matters: Doing Research in the Behavioral and Social Sciences. In S. Greenberg (Ed.), *Human-Computer Interaction: Toward the Year 2000* (pp. 152-169).
10. Moran, T., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., et al. (1997). "I'll Get That Off the Audio": A Case Study of Salvaging Multimedia Meeting Records. In *Proc. of CHI '97*, Atlanta, GA. 202-209.