Evaluation of Mobile Collaboration: Learning from our Mistakes

Melanie Kellar
Kori M. Inkpen
Dave Dearman, Kirstie Hawkey, Vicki Ha, Joe MacInnes, Bonnie Mackay
Mike Nunes, Karen Parker, Derek Reilly, Malcolm Rodgers, Tara Whalen

Technical Report CS-2004-13

September 13, 2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

Evaluation of Mobile Collaboration: Learning from our Mistakes

Melanie Kellar, Kori Inkpen, Dave Dearman, Kirstie Hawkey, Vicki Ha, Joe MacInnes
Bonnie MacKay, Mike Nunes, Karen Parker, Derek Reilly, Malcolm Rodgers, Tara Whalen
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada B3H 1W5
{melanie, inkpen, dearman, hawkey, vha, macinnwj, bmackay, nunes, parker, reilly, rodgers, whalen}
@cs.dal.ca

ABSTRACT

In this paper, we report on the methodological challenges involved in conducting collaborative mobile computing. We investigated two social navigation studies in which pairs of participants were collaborating together over small, mobile, devices. Both of these studies took place in the field, in outdoor and unstructured environments. We reflect on the difficulties identified in previous literatures in addition to new obstacles we encountered. Collaborative mobile computing in the field presents several methodological challenges by combining the problems of many research areas (ubiquitous and mobile computing, and co-located collaboration). However, this avenue of research has many potential benefits to significantly impact real life.

Author Keywords

Ubiquitous Computing, co-located collaboration, CSCW, mobile computing, methodology, evaluation, social navigation.

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Collaborative computing. H.5.3 Group and Organization Interfaces: Evaluation/methodology.

INTRODUCTION

Ubiquitous Computing research and development present a variety of methodological challenges that have yet to be overcome. Abowd and Mynatt [1] address some of these challenges and recognize that the "everyday practices of people" must first be understood in order to be supported, and then must be supported in an integrated, natural manner. It is further stated that novel methodologies are required for conducting human-centric research and developing technologies under this paradigm. This is an exciting time for ubiquitous computing research as methodological approaches are shared and research is allowed and even encouraged to make mistakes. However, in order for this to be worthwhile, reflection upon these experiences is key.

We need to understand the impact of a radically different paradigm on the application of standard field research practices and innovative methodological approaches. On the whole, previous approaches to methodological inquiry in Ubiquitous Computing have been haphazard. It is not enough to say "this is what we did, and it seemed to work". Instead, we need to reflect on *why* something worked and on *how* it could be improved or adapted to other circumstances. Disseminating methodological expertise must include deceptively simple details such as how audio was captured during heavy traffic and how notes were taken during various weather conditions. Results obtained in the field can be even more difficult for other researchers to reproduce simply because details that are simple to control in the lab may only be possible in the field with extreme diligence. We need to be particularly explicit when describing our approaches to research in the field until such time as well established approaches exist.

The goal of this research is to present a detailed discussion of our attempt at conducting collaborative mobile computing research in the field. We discuss the successes and failures we encountered, in order to inform the community about the challenges involved in evaluating collaborative mobile computing. Through this work, we hope to encourage methodological discussion among researchers involved in mobile computing field research.

In general, CSCW research has an added layer of methodological complexity due to the interactions that occur between participants collaborating together. People who are working together can exhibit a wide range of actions, such as gestures, eye contact and verbal cues, which give insight into the nature of the collaboration.

These interactions can often be difficult to capture in the lab and even more difficult in the field. Mobile collaborative computing combines the difficulties presented by ubiquitous computing with those presented by CSCW research.

Many researchers with backgrounds in other HCI-related areas are now considering or conducting what can be termed Ubiquitous Computing research and are moving from controlled laboratory research to field settings. Reporting methodological successes and failures is of critical importance for this group. Our own research background lies predominantly in CSCW, specifically co-located collaboration using various hardware, interaction, and visualization approaches. Part of the motivation for the research discussed here is to explore ways in which our existing research techniques can be adapted to collaborative mobile computing in the field. However, many of these techniques that have been proven to be robust in the lab may fail in the field.

We conducted two complementary social navigation studies exploring ways in which technology can support groups of people when navigating foreign and familiar spaces using maps and other aids. There exists a substantial body of work concerning how individuals use maps and environmental cues to navigate [Cognitive Mapping], providing valuable insight when developing supporting technologies. It is important to realize, however, that reading maps and getting around in public is often done collaboratively. Collaborative use of maps adds an additional social dimension that must be appreciated if we are to develop systems that support navigation.

In this paper, we provide a firsthand account of the challenges and difficulties we encountered during our two studies. Based on our experiences, we present three methodological challenges in addition to the three presented by Abowd and Mynatt [1]. Our first attempts have been learning experiences, and while not pretty, the lessons learned are invaluable for researchers considering taking that first step out “into the wild”.

We first present an overview of the related work examining evaluation methodologies for ubiquitous computing and collaborative mobile computing. We then describe the two social navigation studies we conducted, followed by a discussion of the methodological challenges encountered during these studies. Finally, we end with a brief discussion on the future of evaluation in the field of ubiquitous computing

RELATED WORK

As the field of ubiquitous computing matures, evaluation is becoming a crucial component of all research projects. Previous evaluations in ubiquitous computing research have focused on traditional methodologies [22], exploring ways to apply these techniques to this new genre of HCI research. This section reviews what research methodologies have been used for ubiquitous computing, including the reported successes and failures, and innovations that have been developed to fit the unique challenges of this area.

Evaluation Methodologies for Ubiquitous Computing

In 2003, Kjeldskov and Graham [14] reviewed 102 mobile HCI research papers that were published from 2000-2003. Of these, 42 papers were classified as explicit evaluations. These papers were further classified by the type of method utilized: field studies, lab experiments, and survey research. An additional six papers were classified as case studies. However the motivation behind most of the case studies was system development, as opposed to system evaluation. In follow-up work, Kjeldskov and Stage [16] further examined papers dealing specifically with usability evaluation. They concluded there were very few new methodological approaches being applied during evaluation and within these new approaches there was little variety.

The use of evaluation methodologies is an important topic within the ubiquitous computing community, as evidenced by workshops at Ubicomp 2001 and 2002. Both workshops focused on evaluation methods for ubiquitous computing [20, 21]. The goal of the 2001 workshop was to investigate appropriate evaluation methods for Ubiquitous Computing research, while the 2002 workshop aimed to explore whether or not the approaches being used within the community had evolved to the point where “best known methods” could be established for evaluation of ubiquitous computing systems.

Challenges of Traditional Methods when Applied to Ubiquitous Computing

Applying traditional HCI evaluation methodologies to Ubiquitous Computing research exposes the benefits and pitfalls of these techniques. Overall, research methodologies in this area can be grouped according to setting [22] using Benbasat’s [3] categories of *natural*, *artificial*, and *environment independent* setting.

Evaluation in natural settings includes methodologies such as case studies, field studies, and field experiments. Case studies have been used to better understand complex experiences, but the results can be difficult to generalize [5]. Field studies enable researchers to gather rich observations that are high in realism, however there is no assurance that observations are representative and unbiased [15]. Field experiments afford greater control over the study, helping to ensure that the behaviors of interest are observed [14] and the results are replicable [7]. Although field studies and field experiments are often preferred, data collection and analysis is often the primary deterrent. This was very apparent in Moran et al’s [17] work where the researchers experienced severe difficulty in analyzing audio data. Given the complexity and lack of control within a field study, the value of such studies has been questioned by researchers [15], who compared the results of a laboratory and field experiment. It has been suggested [7] that field experiments represent 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. It is these factors that will undoubtedly impact the eventual use and effectiveness of the system.

Evaluation in an artificial setting involves laboratory experiments. Over 70% of the papers evaluated by Kjeldskov and Graham [14] involved a lab experiment. This type of evaluation has been shown to be beneficial from the perspective of evaluating small pieces of a complex problem (or system) [7, 13]. However, given that context of use is so fundamental to ubiquitous computing, laboratory experiments must be combined with field research methodologies to fully understand the impact these systems have in a normative environment.

Environmentally independent evaluation has consisted mainly of survey research [18] for descriptive data gathering from large samples, but this technique has been used sparingly for mobile ubiquitous computing. [14]

New Methods for Ubiquitous Computing Evaluation

More recently, researchers have attempted to overcome the inherent disadvantages of traditional methodologies and adapt them for use within the domain of ubiquitous computing. For example, Experience Clip [12] is a technique to overcome the intrusive influence of a researcher in the mobile context by encouraging users to take short video clips themselves. A second technique, proposes situative and participative enactment of scenarios (SPES) [10], in order better understand context for ubiquitous computing. This approach 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, other researchers, such as Intille et al. [11], have explored novel tools to facilitate data collection using context-aware experience sampling techniques.

Collaborative Mobile Computing

As previously discussed, evaluating mobile computing in the field can present a variety of challenges. Collaborative mobile computing, whether distributed or co-located, adds an extra layer of complexity. Capturing the rich set of interactions that occurs between collaborators can often be problematic in a controlled laboratory setting and these problems are exaggerated in a field setting.

Distributed collaborative mobile computing is characterized by two or more people collaborating over some distance, with the use of mobile technology. Often participants are in motion and in an uncontrolled environment, such as a downtown area. Because participants are not situated side-by-side, it is not necessary for researchers to capture as many interpersonal interactions occurring between participants, such as eye contact, body language and gestures. However, instead of allocating a researcher to observe each group, as in co-located collaboration, a researcher is needed for each participant if direct observation is required.

Weilenmann [19] conducted a study on the use of mobile awareness devices called Hummingbirds while on a ski trip. She reported that as a single researcher involved in a "field situation [involving] mobile users who are physically distributed … it is simply not possible to be everywhere at the same time". Instead, focus groups and direct observations by the researcher (when nearby) were used. In other work, Grinter and Eldridge [9] conducted a study to examine teenagers' text message behavior. It was noted that it was impractical to conduct direct observations because text messages could be sent from home late at night or while at school. Participants involved in the study also commented that direct observations would inhibit their normal text messaging behavior. In this case, questionnaires, logging data and discussion groups were used. In a laboratory situation where the tasks are contrived, the researchers are aware beforehand who will be engaging in all tasks. In a study such as this one - a field study where participants are following their everyday behavior - researchers are not aware with whom the participants will be text messaging, making it impossible to directly observe all interactions.

Co-located collaborative mobile computing involves two or more people collaborating and communicating side-by-side with the use of mobile technology. As with distributed collaborative mobile computing, the participants may be in motion and in an uncontrolled environment. One of the biggest challenges surrounding co-located collaborative mobile computing is the capture of interpersonal interactions that occur between people who are working together. When co-located collaboration is studied in a laboratory setting, researchers often try to capture interactions such as eye contact, gestures and body language. Many of these interactions can offer clues as to how well the groups are collaborating. Imagine two tourists, standing on a busy street corner, trying to decide on a route. Both of them say "that way" in unison and point in some direction. It is important for the researchers to know if they are pointing in the same direction or in different directions. If captured, these types of interactions can be very useful.

Very little work has been conducted to study how these interactions can be captured in the field. Previous work that has taken place in the field has been primarily in controlled environments. Danesh et al. [6] studied students using handheld devices in order to take part in Geney, a collaborative educational activity. This study took place in a school and the methods of data collection included direct researcher observations and video recording. Bijork et al. [4] conducted a study to evaluate Pirates!, a multiplayer mobile game, where players physically move around the game environment. This study took place at a conference center and data was collected through direct observations, video, questionnaires/interview and system logs. Pairs of participants toured a museum using an electronic guidebook in Grinter et al.'s [8] field study, where data was collected through logging of electronic guidebooks, video recording and participant interviews. Although all of these previous studies were conducted in field situations, they were indoors and in semi-controlled environments, where weather, access to electricity, and network connectivity were not a problem.

In this work, we will present two separate field studies. The first study consists of a social navigation study in which participants were working together in a distributed environment to coordinate a rendezvous. The second study consisted of participants in a co-located situation, working together to complete navigations tasks within a city-wide scavenger hunt.

CHALLENGES IN EVALUATING MOBILE COMPUTING

Two Mobile Collaboration Field Studies

Our current research goals involve the investigation of how people can utilize technology to collaborate in mobile environments. In particular, we are interested in small groups of individuals who have access to mobile devices and need to collaborate to accomplish a task. We will report on two field studies we have conducted to explore issues of mobile collaboration: Social Navigation and City Chase.

Social Navigation

The Social Navigation study explored the activity of rendezvousing. In this study we were investigating the impact that location-aware handheld technology would have on rendezvousing. Working in an outdoor, distributed, downtown environment, pairs of participants completed three navigation scenarios using an electronic map on handheld computers. The handheld map application allowed each user to view the current location of their partner. The three different technology conditions explored in the study were: mobile phones, location-aware handheld devices, and both mobile phones and location-aware handheld devices. The rendezvous scenarios included arranging a rendezvous while separated, negotiating a new rendezvous location when one participant is unresponsive, and awaiting one partner's late arrival to the rendezvous location.

We mimicked a wireless location-aware system using a Wizard of Oz approach that involved two research teams, each made up of two researchers who followed a participant. The two research teams were usually separated by two or three city blocks. They communicated the location of their participant using 2-way radios and updated the participants' current map locations using wireless devices.

For this study, we used four data collection methods: audio, data logging, participant questionnaires and interviews, and researcher notes in the form of the diaries shown below. To capture communication between the pairs of participants, each participant was equipped with an audio recorder. In addition, participants using handheld computers were given HP iPaqs with data logging software installed to capture all interactions with the location-aware system. Finally, participants completed questionnaires after each of the scenarios during and after the study session and took part in an informal interview once the testing had finished.

City Chase

In our City Chase study we explored the benefits of providing shared annotations across multiple mobile devices. In particular, we were interested in comparing users with completely separate views on their handhelds to users who were able to coordinate their independent views with each other. The first phase of this research involved three pairs of participants utilizing handheld computers in a navigation task. City Chase [2] is an organized event in which teams are required to navigate a city, solve clues, find race pit-stops and perform unusual tasks at these stops. The objective of the race is to finish these challenges before any other team. Our three research teams were equipped with handheld devices, paper and electronics maps, and bus schedules to help them achieve this goal.

In this field study we utilized four main data collection methods. The first was audio data. All of our participants were equipped with audio recorders to capture communication that occurred between the participants. The second was video data. A researcher followed one of the teams throughout the day, capturing video of interesting navigation exchanges between the participants. The third data collection method was data logging on the iPaq handheld computers. Finally, participants took part in a verbal debriefing session with the other researchers following the race and wrote retrospective discussions of the day's events (also leading to the diaries below). Although these seem like very basic, routine data collection methods, within the realm of evaluating collaborative mobile computing, nothing is really that simple.

The initial goals of these studies were to gather some insights on our research question; pilot the software and hardware and better determine what functionality would be needed; and to explore different methodological approaches for data collection.

Ubicomp in the Field: Social Navigation

'Do you want to meet for coffee this afternoon? Sure, what time? Where?' This social act of negotiating a meeting place and time occurs daily, as does the need to re-negotiate a meeting place and time due to unexpected happenings, such as traffic or a meeting that runs late. If people have cell phones they can quickly call to modify existing plans. Those who have handheld computers with location-aware software can visually view the exact location of their family members/friends/colleagues and propose an alternative meeting spot and time that is conveniently displayed for all involved. The main research focus of this study was to explore the impact of location-aware handheld technology on the act of rendezvousing. Secondly, we were interested

in examining how to collect meaningful data while performing a field study using collaborative mobile technology in a realistic setting with a variety of uncontrollable variables.

July 30th, 2004 – Day one

After having participants use the cell phones only, and location-aware software on the PDA's only, we are now testing the combination of cell phones and PDA's. The two participants finally show up at the rendezvous location for the study. Interestingly, the simple act of meeting with study participants to do the study requires a successful rendezvous! The participants read and sign the consent forms and then are given a quick training session on the location-aware software and also on the cell phones. We set off – despite the odd looks from other pedestrians. For each participant, there is a team of two researchers: one researcher is talking with the other team to track the location of the other participant, and the other researcher is taking notes. Each participant is given a different task to complete and instructions on how to set up the rendezvous location. While participants have been told to obey traffic lights and not to run (so the research teams can keep up), participants still want to rush to beat lights and are weaving in and out of the crowds. Keeping notes and the locations on the PDA's up to date can be challenging, especially when trying to dodge the crowds on the sidewalk. Still, we keep up and for the most part maintain the illusion of GPS enabled devices. The participants use their handhelds to check the progress of their teammate and also use cell phones for quick updates. I wonder if it easier for the participants to arrange a new meeting place using the software instead of trying talk over all the noise on the street on the cells. It will be interesting to find out what the participants think of using the location-aware software and the cell phones after we finish all the scenarios.

Ubicomp in the Extreme: City Chase

Imagine running through the city, hopping busses, and riding ferries. Mix in a time pressure component and various "challenges" and you have an exciting navigation adventure. Instead of a contrived scenario with artificial tasks we wanted to explore the use of ubiquitous computing 'in the wild'. The City Chase was a local urban scavenger hunt that provided a unique test bed to explore collaborative navigation using mobile devices. Although a scavenger hunt isn't a normal day-to-day activity, it did provide a scenario where participants' motivation to navigate efficiently was of critical importance. Our primary research goal for this work was to investigate technology to maintain coordinated views across devices in order to support mobile co-located collaboration. As with Social Navigation, our secondary research goal was to explore various methods for collecting data from pairs of mobile participants collaborating over devices. It is this second research goal that we present in this paper.

July 15th, 2004 (three weeks before the City Chase)

We are running in the University Athletic Facility, testing out the most recent version of our software and trying to find a voice recorder and microphone setup that will be unobtrusive yet enable the capture of reasonable audio data. Today we are testing the gear setup (backpacks, headsets, shoulder-belts and fanny packs) to see if we can still run around carrying all of this equipment. Just imagine a Cyborg triathlete. Although we get a good sense of what works and what doesn't, this unfortunately isn't a realistic test because we are running inside an athletic dome. Many of the environmental factors are more controlled in this setting, including background noise and lighting. However, it's pouring rain outside right now and we haven't yet figured out how to protect our gear from getting wet. We hope it doesn't rain the day of the event!

August 7th, 2004 – Race Day

Guess what? It's raining! Not just drizzling, absolutely pouring! We casually stuff our iPaq 4100s into plastic baggies, quickly laminate our paper maps and hope that the cell phones are waterproof. I guess this constitutes real usage – it isn't practical to only be 'sunny-weather researchers'. The first leg of the race is a scavenger hunt to find 10 out of 15 items on a list. No need for technology yet - just run around and scrounge up a piece of rope, some salt water taffy, a Sou' Wester rain hat and a variety of other odds and ends.

As the day progresses we make our way around the city, trying to find the most optimal route, traveling on buses and ferries, and conquering various challenges such as eating a can of sardines and a live worm (all in the name of research). Throughout the day we utilize various media for assistance, such as cell phones, handhelds with map navigation software, paper maps, transit schedules, and paper and pens for note-taking. Four hours and eight minutes later we cross the finish line, exhausted. Our other two research teams are not far behind.

METHODOLIGICAL CHALLENGES

The methodological challenges we encountered during our two studies covered a wide range issues. The first three challenges we reflect upon were previously identified by Abowd and Mynatt [1] in March 2000. Despite the many changes that have occurred in the ubicomp community since this time, these issues are still problematic today. Following this, we present three additional challenges related to data collection, particularly relevant to the evaluation of collaborative mobile computing.

#1 Finding a Human Need

The ease of building ubiquitous computing applications has improved significantly since the writing of Abowd and Mynatt's [1] paper. However, the requirement to satisfy a real or perceived human need is still an important issue. In many cases, the ubiquitous applications that we want to study are not widely deployed and as such it is still difficult to evaluate "the impact of a system on the everyday life of its intended population" [1]. In the Social Navigation study, the navigational tasks represented real-life scenarios common to everyday social interactions. In the City Chase, the task of navigating the streets was realistic. However, the City Chase itself was not an everyday task in that it was a timed competition requiring the participants to make quick and calculated decisions on the run. In this study, while we had perceived a use for shared annotations on handheld devices, the participants soon realized that cell phones and paper maps were easier and faster to use. Although the novelty effect of a new technology may temporarily increase the use of technologically innovative solutions it was not so in this case. The pressures and pace of the race event actually dissuaded participants from the handheld solution, and led them to choose the paper map and cell phone combination.

#2 Evaluation in the Context of Authentic Use

Abowd and Mynatt [1] describe the importance of deployment of ubiquitous computing into authentic settings. One important reason for this surrounds motivation. In artificial environments, tasks are contrived and participants' motivations to complete the tasks (and in many cases, utilize the technology) are artificial. Also, it can be difficult to generalize results to real world situations when the tasks are artificial. In our research, we were looking for scenarios where participants would utilize technology to facilitate a need outside of the research study itself.

The Wizard of Oz methodology for the Social Navigation study allowed us to study participants' real behaviors during our experimental rendezvous on the busy streets in Halifax. Still, while the Wizard of Oz approach allowed participants to actively engage in real rendezvousing behavior with their partner, it also influenced the normal actions each participant would have made had they not been followed by the team of researchers. Firstly, to accommodate both the research teams following each participant and to enable data collection, the participants could not dash off, run or try to 'beat traffic' as they might in real life. Secondly, the constant presence of the researchers meant that participants were frequently asking them questions regarding the task at hand.

In the City Chase, we overcame this challenge of authentic settings by taking advantage of a 'scavenger hunt' which members from our research group were participating in and provided mobile computing support for those participants. While not a true, everyday, potential usage of mobile technology, the scavenger hunt provided external goals for the participants, outside of the requirements imposed by the study itself. The participants' primary goal was the race itself and we hoped that this would provide realistic usage. The fact that we provided an authentic situation where the participants' motivations were an external goal resulted in an important outcome – the participants essentially did not utilize the technology. Each group only utilized the technology once for navigation purposes (the intended usage of the device) while one group used it one additional time for a non-navigation task. Although the usage context was real, the use of small, mobile devices for navigation was not a real usage for our participants. Without a strong level of familiarity and comfort with these activities, it was easier to rely on more traditional means for navigation (i.e. call a friend).

#3 Task-centric Evaluation

To properly evaluate ubiquitous computing applications, we need to evaluate systems using tasks that are truly reflective of the actual need driving the application. If you do not know or distinguish all the tasks that will be performed, then it makes if difficult to identify and apply the appropriate evaluation methods.

In the Social Navigation study we chose tasks that ranged from simple tasks, such as doing an errand and then negotiating a nearby meeting place with the other partner, to more complex tasks, such as negotiating a rendezvous when one partner is delayed. We also examined how participants dealt with technology failure within a task, for instance, we simulated cell phones that were out of range and handhelds that didn't have wireless connectivity. We did not evaluate was the effect of familiarity with the city on rendezvousing tasks. Almost all of our participants were familiar with the downtown city core and we expect that if one or both participants were unfamiliar with the area the rendezvousing behavior would be affected.

In the City Chase, we made educational guesses as to the different tasks that would be performed while doing the race based on a practice run and used this information to help determine our evaluation methods. We did not anticipate how the task would be influenced by the pressure of the race and outside factors, such as rain. We also did not know the tasks that would be performed so it was difficult to identify appropriate evaluation methods. We guessed at many potential tasks, but guessed wrong in many cases.

#4 Collection of Verbal Data

Given our interest in how people collaborate when using small, mobile devices, it was important to record conversations that occurred between the participants. Recording good quality audio is very difficult in the field, particularly when the participants are mobile. We chose to outfit the participants in both studies with voice activated audio recorders equipped with automatic transcription software.

The first obstacle to overcome was how the participants would carry the audio recorders. In the Social Navigation study, participants were given pocket-sized voice recorders. Most participants carried them in their jacket pocket and small microphones were clipped to their lapel. In the City Chase, the voice recorders were more problematic due the fact that our participants were running between locations and would have many other objects in their hands (e.g. handheld computers, maps). We purchased knapsacks from a local hiking store and ran a microphone from inside the knapsack out to where it could be clipped on the shoulder strap (where it would pick up audio). Figure 1 shows a picture of the knapsacks. Overall, carrying the devices in the knapsacks worked well with the exception that in one case the device accidentally turned off inside the knapsack and there was no visual indicator that audio was no longer being recorded. Similarly, in the Social Navigation study periodic checks had to be made to ensure that the audio recorders were still capturing voice data intruding on the naturalness of the participant's scenarios.

Figure 1. Knapsacks used to store audio recorders and other supplies needed the day of the race.

The second obstacle central to both studies was how well the devices would operate in the field. Environmental noise, such as busy street traffic and pedestrian conversations, was problematic in terms of the functioning of the device as well as the quality of the audio received. For example, in the City Chase the devices were equipped with a voice activation feature to minimize storage requirements and battery consumption. However, the sheer ambient noise present outside activated the devices so that they never shut off. In addition, the ambient noise obscured the participants' voices on the recording, making it difficult to clearly record the conversations. In the City Chase, participants were co-located, so while outside noise interfered with the logging of interactions, both sides of the conversations were on one audio recorder. In the Social Navigation study, analyzing the audio recordings was more of a challenge. Participants were separated with each wearing their own audio recorder, so before we could analyze the data we had to first converge the individual recordings into one legible conversation. Furthermore, the automatic transcription software wouldn't work given the poor quality of the audio. An excerpt of the audio from the City Chase study audio can be heard at: http://www.edgelab.ca/citychase/audio.

#5 Collection of Video Data

Beyond audio data, it would also be beneficial to have a record of the participants' behaviors while they were attempting to collaborate on a task. As was the case for audio data, collection of video data is problematic when the participants are mobile. While video capture would have been advantageous in capturing robust information in the Social Navigation study, participants were already accommodating a team of two researchers, making the addition of a video camera more distracting for the participants. As well, we felt that we could not ensure quality video on the busy downtown Halifax streets. In the City Chase, this problem of using video was exacerbated given that the participants typically ran between locations. We attempted to have an observer follow each pair and record video at various points throughout the activity. A few days before the event, we piloted the feasibility of our running observers. Of the three observers (one per team), only one was able to physically keep up with their team. We discussed the possibility of having the observers drive between locations but this was problematic given parking, driving routes, and the fact that we didn't know the locations ahead of time. We also discussed the possibility of an alternative form of transportation (e.g. bike or roller blades) but recognized the difficult of handling a video camera while biking or roller blading around the city. Finally, it was decided that only one person would attempt to follow a team around and collect video data (see Figure 2).

Figure 2. The observer (on the left) followed the team with a video recorder during the City Chase.

We refer to our video collection technique as the "Blair Witch Method"[1] given the quality of footage we received.

[1] The Blair Witch Project is a video produced by Artisan Entertainment, released on DVD on August 21, 2001. This video is a mock documentary of students investigated a local urban legend (The Blair Witch). One of the key distinguishing aspects of this movie was its poor video quality, the kind that you would get

Video recording outside, in a public location, with participants that are constantly moving around does not represent an ideal situation. Overall, the video data is shaky and the viewing angles are less than ideal (see Figure 3), however, it does provide data that we otherwise wouldn't have. An excerpt of the video data can be viewed at: http://www.edgelab.ca/citychase/video.

Figure 3. A screenshot of the shaky video taken during the City Chase.

#6 Environmental Constraints
In a lab study, researchers can control for variables such as lighting, temperature and noise. In field studies you can control, to a certain extent, the tasks your participants undertake, but you must be prepared for uncontrollable factors such as traffic, weather, and outsiders' actions. Perhaps the most uncontrollable is the weather. Rain caused the delay of sessions during the Social Navigation study, and the morning of the City Chase it poured rain (see Figure 4). The cell phones were wet, the paper we were writing on disintegrated, and we had to put our iPaqs in plastic baggies (see Figure 4). The rain was also problematic for our video observer in the race, who could be seen running down the street with his video camera in a plastic bag, trying to shield the rain with an umbrella.

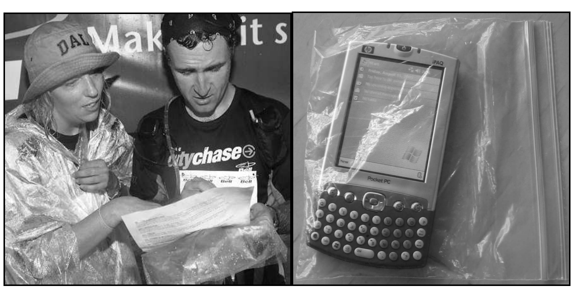

Figure 4. The left-hand picture shows a team working in the rain and the right-hand picture shows how the iPaqs were protected from the rain.

from amateur filmmakers running through the woods with a video camera.

DISCUSSION AND CONCLUSION

Abowd and Mynatt [1] outlined several evaluation challenges for Ubiquitous Computing. These challenges are still evident although advances in ubiquitous computing infrastructures are making some of them more manageable. Despite the caution that formative and summative evaluations are difficult, it is still important to conduct this type of research. There is a "chicken and the egg" issue at work here, in that we can't conduct robust field studies without solid methodologies, yet we can't develop these methodologies without learning from previous research experiences.

Our investigations centered on how people can utilize technology to collaborate in mobile environments. In particular, we looked at small groups of individuals who were collaborating across handheld devices in order to accomplish a task.

The Social Navigation study explored the activity of rendezvousing in which pairs of participants performed navigation tasks using handheld computers to view the location of their partner. The goals of this project were to explore the behavior effect of introducing location-awareness into rendezvousing, as well as exploring methodological approaches for data collection in a mobile collaborative environment.

In our City Chase study we were interested in the benefits of providing shared views across multiple mobile devices. The first phase of this research involved three pairs of participants utilizing handheld computers in a navigation task. The goal of this initial phase was to collect some inital impressions about our research question; pilot the software and hardware and examine what functionality would be needed; and to explore the available methodological approaches for data collection.

We have demonstrated a number of evaluation methods attempts within these two field studies which explore ubiquitous and collaborative, mobile computing solutions for navigation. Although we based our evaluation methodologies on suggestions from previous literature in similar fields, we still had difficulties due to uncertainties in task and environment. It is clear that more work needs to be done (and reported) on successes and failures in evaluation methodology in order to learn from our mistakes. It is inevitable that methodological inquiry in Ubiquitous Computing will stabilize. If we take a page from other research areas, we can see how opportunity for novel methodology is short-lived. We should be keenly aware of this and foster reflective methodology as a key aspect to all future ubicomp research.

ACKNOWLEDGMENTS

Thanks to Dave Parsons and Haewon Chung for their participation in this project. This research was funded by NSERC.

REFERENCES

1. Abowd, G. and Mynatt, E. (2000). Charting Past, Present and Future Research in Ubiquitous Computing. *ACM Transactions on Computer-Human Interaction, Special issue on HCI in the new Millennium, 7*(1): 29-58.
2. *The Bell City Chase*. (2004). from http://www.thecitychase.com/en_index.asp
3. Benbasat, I. (1985). An Analysis of Research Methodologies. In *The Information System Research Challenge* (pp. 47-85). Boston, MA: Harvard Business School Press.
4. Bjork, S., Falk, J., Hansson, R., and Ljungstrand, P. Pirates! Using the Physical World as a Game Board. In *Proceedings of.*
5. Cheverst, K., Davies, N., Mitchell, K., and Efstratiou, C. (2001). Using Context as a Crystal Ball: Rewards and Pitfalls. *Personal and Ubiquitous Computing, 5*(1): 8-11.
6. Danesh, A., Inkpen, K. M., Lau, F., Shu, K., and Booth, K. (2001). Geney: Designing a Collaborative Activity for the Palm Handheld Computer. In *Proceedings of SIGCHI '01*, Seattle, WA. 388-395.
7. Goodman, J., Brewster, S., and Gray, P. (2004). Using Field Experiments to Evaluate Mobile Guides. In *Proceedings of the 3rd Annual Workshop on HCI in Mobile Guides*, Glasgow, Scotland.
8. Grinter, G. E., Aoki, P. M., Hurst, A., Szymanski, M. H., Thornton, J. D., and Woodruff, A. (2002). Revisting the Visit: Understanding How Technology Can Shape the Museum Visit. In *Proceedings of CSCW '02*, New Orleans, LA. 146-155.
9. Grinter, G. E. and Eldridge, M. A. (2001). Y Do Tngrs Luv 2 Txt Msg? In *Proceedings of ECSCW '01*, Bonn, Germany. 219-239.
10. 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.
11. Intille, S. S., Munguia Tapia, E., Rondoni, J., Beaudin, J., Kukla, C., Agarwal, S., et al. (2003). For Studying Behavior and Technology in Natural Settings. In *Proceedings of UbiComp 2003*, Seattle, WA. 157-174.
12. Isomursu, M., Kuutti, K., and Väinämö, S. (2004). Experience Clip: Method for User Participation and Evaluation of Mobile Concepts. In *Proceedings of the Eighth Conference on Participatory Design*, Toronto, Canada.
13. James, C. L. and Reischel, K. (2001). Text Input for Mobile Devices: Comparing Model Prediction to Actual Performance. In *Proceedings of SIGCHI '01*, Seattle, WA.
14. Kjeldskov, J. and Graham, K. (2003). A Review of Mobile Hci Research Methods. In *Proceedings of Mobile HCI 2003*, Udine, Italy.
15. 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 *Proceedings of Mobile HCI 2004*, Glasgow, Scotland.
16. Kjeldskov, J. and Stage, J. (2004). New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human Computer Studies, 60*(4-5): 599-620.
17. 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 *Proceedings of SIGCHI 1997*, Atlanta, GA. 202-209.
18. Schmidt, A., Stuhr, T., and Gellersen, H. (2001). Context-Phonebook - Extending Mobile Phone Applications with Context. In *Proceedings of Mobile HCI 2001*, Lille, France.
19. Weilenmann, A. (2001). Negotiating Use: Making Sense of Mobile Technology. *Personal and Ubiquitous Computing, 5*: 137-145.
20. *Workshop on Evaluation Methodologies for Ubiquitous Computing*. (2001). Retrieved September 10, 2004, from http://zing.ncsl.nist.gov/ubicomp01/
21. *Workshop on User-Centered Evaluation of Ubiquitous Computing Application*. (2002). Retrieved September 10, 2004, from http://zing.ncsl.nist.gov/ubicomp02/
22. Wynekoop, J. L. and Conger, S. A. (1990). A Review of Computer Aided Software Engineering Research Methods. In H.-E. Nissen, H.K. Klein and R. Hirchheim (Eds.), *Information Systems Research: Contemporary Approaches and Emergent Traditions* (pp. 301-325): Elsevier Science Publishers.