USABILITY TESTING

Google Maps for Mobile

SI622: Evaluation of Systems and Services
March 28, 2008

Jacqueline Cerretani
Beilei Zhang
Jim Laing
Gaurav Anand

smellyphone@ctools.umich.edu

1. INTRODUCTION

This document describes the user tests of Google Maps on Mobile (GMM) application as experienced on Blackberry Curve 8310 mobile phone. These tests contribute to more rigorous, semester-long student project for the course SI622: Evaluation of System and Services, carried out under the guidance of Prof. Mark Newman. The goals of these tests are to evaluate the usability of the application, determine prioritized issues for improvement, and propose possible solutions to the developers. This report will assist the professionals responsible for the development and marketing of future versions of the application.

1.1 About Google Maps for Mobile

Described as "the power of Google Maps on a mobile phone," Google Maps for Mobile (GMM) delivers step-by-step driving directions, traffic information, satellite imagery, and interactive mapping on mobile phones. The application takes advantage of Google's informational prowess by allowing users to search for map features by category, vague location or colloquial name, and returns results that include location, phone number, reviews and relevant details. GMM integrates search results with phone services via its click-to-dial feature, and allows users to personalize the application with an option to save favorite locations. In addition, the latest version of GMM is location aware, allowing users to map to and from their present location.

1.2 Target population for GMM

Though we do not have an explicitly stated target audience for GMM from its development team, an assessment of market trends and user data we've collected thus far has revealed a few potential characteristics. As a priori characteristics, we postulate that GMM is targeted toward users with a moderate to high level of mobile device savvy, since the application runs solely on data-enabled smart phones, and secondly, GMM is targeted towards users of Google Maps Online (GMO), since the application is designed in many ways to mimic the functionality and features of the online version. This is bolstered by the fact that our survey revealed that most users of GMM are also users of GMO.

We also postulate that GMM is pursuing two emerging market segments. First, we believe they may be attempting to capitalize on the exploding population of mobile phone users who are converting to data-enabled smart phones. In recent years, smart phone sales have been growing enormously; between October 2005 and October 2006, smart phone sales increased by 75%. There has also been a complimentary growth in the development and adoption of applications for smart phones: in 2005, 11,000 new smart phone applications were released. This indicates that more and more people possess smart phones and are becoming proficient with applications on them, providing GMM with a potential conversion population that is ripe for extending their desktop experience with Google Maps to a version on their mobile phone. This postulation is bolstered by the fact that the GMM team appears to be actively working toward extending the application to most smart phone brands.

In addition, judging from their recent launch of geographic awareness in November, 2007 (through cell phone tower triangulation technology), we postulate that they are actively working to become a competitor to other mobile mapping services, such as TeleNav, VZNavigator, and Nokia Maps, and possibly even to hand-held GPS-only units. In our interviews, competitive analysis and surveys, we found that most GMM users have at one time used handheld GPS devices. In 2006, Gartner postulated that over 40% of handheld devices will be GPS enabled in just two years. With GPS enabled smart phones becoming increasingly popular and the development of a complimentary technology for those which are not, GMM can capitalize on this potential conversion group, who may find the convenience of high-level mapping capabilities unified with their day-to-day mobile device appealing.

2. USER TEST GOALS

The broad goal of user tests is to prepare a prioritized list of usability problems that are not discovered by heuristic evaluation, along with suggestions for improvement that can be use by the development team of the application. More specifically, we targeted for following sub-goals:

- To uncover and document hidden usability issues that experienced BlackBerry users might encounter when first learning to use GMM.
- To uncover and document user interface bugs that veteran users of GMM are likely to encounter on a frequent basis.
- To gain a better understanding of the usage patterns likely to be exhibited by GMM users.
- To research and gain experience in conducting mobile usability testing on a limited budget.

3. TEST METHODOLOGY

As the "gold standard" of system evaluation, user testing is designed to uncover insidious usability issues that would not necessarily be identified in other forms of evaluation. In order to take advantage of the unique traits of user testing, we decided to focus on the evaluation of common tasks that GMM users are likely to undertake. By observing users performing such tasks, we were able to collect data on the efficiency, accuracy, and emotional response of our subjects (Schusterisch et al. 2007). In doing so, we uncovered common points of frustration and confusion.

We conducted seven user tests, including two pilot tests, in order to refine both our testing platform and protocol and to identify a broad sampling of usability issues. This number is consistent with Jakob Nielsen's research indicating that five usability tests is the ideal number for identifying most issues since several of our tests where slightly hindered by our equipment research (Froehlich et al. 2007).

We used the BlackBerry Curve 8300 as our standardized test platform. As our survey revealed, BlackBerry is a very common device used for running GMM, so we expect that our findings using this as our test platform will be representative of the usability issues encountered by many users of GMM. We utilized recognized and respected user testing methods, including a standardized protocol for interacting with test subjects, audio and video recording using a speak-aloud, and follow-up with post-test interview.

3.1 Task Design and Subject Selection

Recognizing that we only had a limited amount of time and resources to conduct user testing, we decided that we would be most productive if we focused on the common tasks and features in GMM undertaken by BlackBerry users. We based our understanding of common GMM tasks on our survey results, thus removing any bias toward specific features our team found compelling or important. Furthermore, we decided that focusing on current or previous users of

BlackBerry devices would remove the necessity to train our test subjects in how to use the BlackBerry. This would enable us to focus on testing the GMM application, rather than the BlackBerry device or our ability to instruct subjects in its operation (Refer Appendix 8.2).

Our initial approach to developing tasks for user testing was to focus on issues that we uncovered during our previous heuristic evaluation, as a means of confirming that these represented actual usability issues. However, after reflecting on the strengths and weaknesses of both heuristic evaluations and usability testing, we decided to revise this approach. Since heuristic evaluations claim to cover a large portion of the usability issues that could be uncovered by usability testing, we decided to design our tasks in a general way that would hopefully allow us to uncover issues not identified in our heuristic evaluation. In this way, we hoped to extend our understanding on the potential usability issues in GMM.

Finally, we decided to design our tasks based on a possible scenario that the subject could potentially encounter in real life. We did this by asking test subjects to imagine themselves in a situation in which they would be using GMM to plan a meet up with friends and then asking them to walk the route to the meet up destination. By basing our tasks on a real life situation, we hoped to make the tasks more relevant to the subjects while also encouraging them to behave in a more natural way that would better mimic their actual experiences with GMM.

Table 1: User Tasks

User Tasks		
Warm-up	1.	Search for 635 Fifth St., in Ann Arbor, MI.
Actual	2.a	You and your friends are planning on going out tonight for dinner at the New York Pizza Depot on East William St. You've never been, so you decide to look it up on your phone. Tell us when you've found it.
	2.b	Now save it as a favorite in Google Maps for Mobile to use for later. Please let us know when you've succeeded in this task.
	3	You know you'll need to find parking when you go, so you decide to look around the area of the restaurant to see what might be nearby. Pan around to find what looks like the nearest parking garage. Please let us know when you've succeeded in this task.

3.2 Recruiting and Delivery

While showing GMM on the Blackberry to friends who had used it on other devices, we found that they were adroit with the software but had trouble learning the Blackberry's specialized keys. As a result, we chose to recruit only people who had used a Blackberry device at some point in the past. This ensured the greatest likelihood of problems uncovered in our user tests being related to the software rather than the device.

Users were recruited in two ways: first, we called directly upon a pool of volunteers who had asked to participate in the study early in the semester. We also canvassed the SI student population via the si.all email list. We considered this a strategic, rather than convenient, move, as the technological savvy of this group matches that of GMM's target audience. Also, the prevalence of smart phone users ensured that a greater-than-usual portion of this group would have experience with the Blackberry, something we feared might be difficult to come by in a lay audience.

As a result of difficulties we'd had with using a strong qualitative appeal in our last recruitment message, we toned down our recruitment message, appealing a little more to self-interest. We offered a $10 gift certificate to Amazon.com, and highlighted the chance to play with the new Blackberry Curve.

3.3 Test Protocol

We designed our standard user test protocol in order to put our test subjects at ease throughout the duration of the test so that we could observe their natural behavior. After introducing ourselves and our project to each subject, we explained our confidentiality agreement and asked that they sign a consent form (Refer Appendix 8.1), making sure that any outstanding questions they had were answered. We also asked each user to complete a brief survey (Refer Appendix 8.3) indicating their experience with the BlackBerry, GMM, and similar products as well as provide additional demographic information. Finally, we asked that they speak aloud their thoughts and intentions as they interacted with GMM so that we could obtain an accurate picture of their mental processes.

Throughout our testing we also discovered that we needed to improve our orientation to the device. Users had experience with various versions of BlackBerry, but not necessarily the BlackBerry Curve. In our first test we did no orientation to the device; this was very confusing. In the second pilot, we showed the user three keys (the back button, the scroll wheel, and the Blackberry key) during the warm-up test, but this later proved to be insufficient. We revised our protocol to show the following keys, which were all used in the completion of the tasks: Blackberry, scroll button, back, return, delete, shift (for numbers). We also did this before starting the warm-up task and asked them to confirm that they felt comfortable with the interface. This made the task much more helpful.

3.4 Test Platform

3.4.1 Challenges of Mobile Testing

Usability testing a mobile mapping application is a difficult task. Features such as finding one's current location by surveying context and landmarks, and following a route on foot, are impossible to test indoors or standing in one place. Other features, such as following a route in a car, are almost unethical to test, due to danger of the mounting distractions of using the device, talking aloud, being recorded and being questioned. In addition, few of the conventions or environments useful for testing a desktop application apply in this field. Video of facial expressions, audio and screen capture cannot all occur on the mobile device due to a lack of available accessories as well as memory and CPU limitations. In addition, it would be valuable to capture interaction with the keypad to track down issues related to key mappings. The usability testing environment must be light and portable. Notes taken by hand while walking can turn in to indecipherable scribbles, so they must be done post hoc, or in some other way.

On an unlimited budget and in a heavily technologically enhanced environment, these constraints might be easy to overcome. But in our case, on the shoestring budget of graduate student, creativity had to replace cash. The task is an inspiring one, in that developing a system to conduct usability tests on mobile devices on a shoestring could lead to independent developers adopting the practice and producing much improved applications.

In order to generate ideas and vet our own, we did a limited literature search on mobile usability testing environments. Researchers at Google attempted several iterations of a mobile usability testing set up (Isomursu et al. 2004) first starting by mounting the mobile device, along with two cameras -- one to capture the device screen and one to capture the keypad -- to a piece of wood that the user would hold. However, they found that this setup was too heavy to be used for very long, and interfered with the users' typical typing patterns. This led to the design of a set of clamping devices that mounted the two above cameras on a wide variety of types of phones, and allowed for a third camera to capture the user's facial expressions. The video was then fed to a laptop computer via USB to a software application that could sync all the input feeds, and allow live observation by distributed teams.

While Google's set up looked most promising as a basic framework to emulate, we looked into some alternate, less conventional, methods as well. We examined Jon Froehlich's myExperience method, which prompts users to reflect on their experience with their mobile device in situ at variable times, triggered by a set of interacting sensors, triggers and actions. While this method is interesting, we found it both too complex and the range of actions captured too broad to fit our needs, time frame and budget.

Finally, we considered eschewing all typical data capture methods in favor of a more user-directed system such as that outlined in the Experience Clip method. Isomursu and colleagues conducted an experiment in usability analysis of location aware mobile applications to be used outdoors. Users were paired together and each given a mobile device. One was charged with completing a particular task, and the other was charged with documenting their partner's experience in short video clips. The researchers found that liberating the subjects produced more variable, expressive, and realistic feeling video, and made them more apt to solve problems independently, they also found that subjects were reluctant to let their failures be recorded, omitting critical data from tests. While this environment seemed enticing for its uniqueness, we felt that it was unlikely we could compensate for the variability in quality of data with our small sample size. Furthermore, we anticipated that technical problems might undermine the users' ability to capture. Judging from our subsequent experience, this point was quite prescient.

3.4.2 Equipment

We began the process of designing shoestring mobile test environment by identifying the hardware and software options available to us. The most basic requirement for user testing is a device running GMM. For this, we had access to both a BlackBerry Curve 8310 with GPS location and a BlackBerry Curve simulator from Research In Motion. The BlackBerry simulator can be run on any Windows-based computer, and we had access to several laptops that met this requirement, some featuring built-in cameras and microphones and others without.

Through the course of our testing we also identified several other pieces of hardware which we had ready access to, and which proved useful in our testing experiments. One of these devices was a compact digital audio recorder with microphone which we used to capture audio data during several of our user tests. In later tests we also made use of a headset-style microphone plugged directly into a laptop computer for recording audio. We also had access to a basic external USB webcam, which was fortuitously mounted on a plastic tension clip, allowing us the option of mounting this camera on either the laptop or the BlackBerry device.

Figure 1: Our equipment set (laptop with built-in camera, BlackBerry with mounted camera, and headset)

We also had access to two software tools which provided the data collection tools necessary to supplement our handwritten notes and audio recordings. One of these tools was Camtasia from TechSmith, which is available as a free 30 day trial. This software allowed us to capture video directly from the laptop's screen, as well as simultaneously capture video from an external or built-in web cam and audio from an external or internal microphone. In addition, we used a freeware application called BBScreenStream, written by an anonymous programmer and available online at http://oppitronic.net/pb/bbscreenstream.php. This application captures a screenshot from a BlackBerry device connected via USB cable at designated intervals ranging from every 2 seconds to over 0.5 seconds. Using this application allowed us to capture the BlackBerry device screen exactly as it appeared to the user throughout the duration of the test, significantly improving our capacity to identify meaningful results.

Finally, in order to transport all of the necessary equipment, we used an over-the-shoulder messenger bag, designed to carry a laptop computer. Using this bag allowed us to utilize a laptop computer during all phases of the test. More specifically, it enabled us to position the computer in close proximity to the test subject (which is necessary for cables to reach) while ensuring that the equipment did not interfere with the subject's ability to perform the tasks naturally.

3.4.3 Iterative Refinement

Starting with our two pilot surveys, we went through several iterations and refinements to our test platform in order to best capture the experience of our test subjects. In our first pilot test, we opted to begin the test using the BlackBerry Simulator and then switch to the actual BlackBerry device, conducting both an inside and outside portion of the test. This was done for several reasons. First, we believed that using the simulator would enable us to capture richer data about the user experience than we could outdoors. This is because we were able to capture a higher frame rate screen video, video of the user's face, and clearer audio than we could record outside.

However, we soon discovered that our pilot user had trouble understanding and interacting with the simulator. We believe this is because the presentation on the device on a laptop computer created conflicting mental models as to how to interact with the device. While interacting with the simulator must be done in roughly the same way at the actual device by clicking on individual keys and using the scroll ball, often the user attempted to interact with the device as if it were a typical desktop application. This was most clearly seen when the user would click directly on the screen area of the simulator and use the mouse to select options from the GMM menu, which are actions that the simulator does not recognize. Ultimately this caused the user to spend more time trying to overcome their instincts to use the computer interface and less time on the task. Further tests confirmed that the search task on the simulator was perceived as more difficult than when it was done on the actual device.

For the second pilot test, we decided to abandon the use of the simulator and instead conduct the entire test using the BlackBerry device. However, we still intended to perform an indoor and outdoor portion of the test because we believed it would be difficult or impossible to capture the subject's facial expressions during the outside portion of the test. It just so happened that before this test, we decided to switch the laptop computer we would be using to conduct the test. The replacement laptop was not equipped with a built-in webcam, and so we had to utilize an external webcam, the USB camera with clip previously mentioned. Upon arriving to the user test and setting up our test laptop, our test subject remarked about how the USB camera could be attached to the BlackBerry device to capture video throughout the duration of the test. After considering the matter, we agreed that this was feasible, and decided to capture video of the subject's facial expressions both indoors and outdoors.

After this test, we made several small refinements throughout the remaining tests, including switching from the use of an external audio recorder (a vestige of our earlier tests using two separate capture methods) to using a headset-style microphone capturing audio through the laptop's microphone port. This increased the clarity of the audio and eliminated the need for us to synchronize the audio file with the video capture file.

Figure 2: Blackberry handset with mounted camera

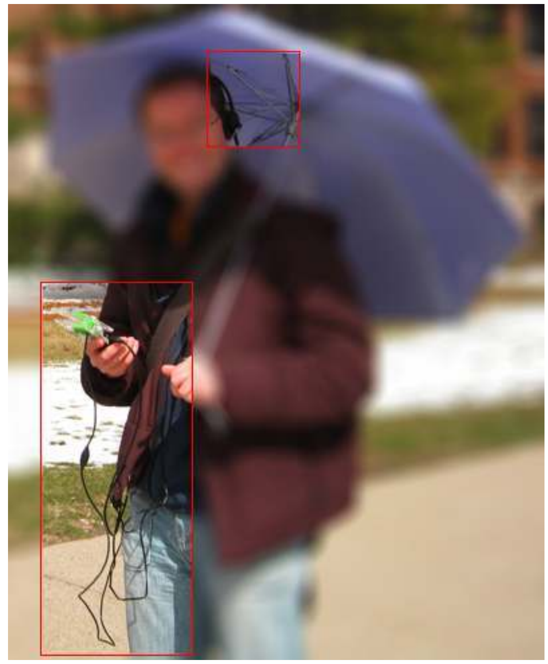


Figure 3: One of the users wearing full equipment gear on bright sunny day

We also decided to move the entirety of the test to the outside. This is because we found that the tasks performed inside took much longer than they did outside, because of the poor cellular communication in various institutional buildings. Switching environments also caused a disjunction between the tests for the users; though we cannot say our data displays this difference, we noticed that users completed the tasks much more slowly for both, and didn't seem to do them with as much facility, even though components of the first task were reused in the second tasks.

3.4.4 Unexpected Challenges

Despite our two pilot tests and our continuous refinement of our user testing platform, we still discovered many unexpected challenges that demonstrate the difficulty of mobile user testing. We experienced several technical problems, due the complexity of our testing platform, as well as various environmental interferences that interfered with our ability to collect all of that data available.

Our technical problems included misconfigured laptops, loose cables, buggy software, signal interference, and audio noise. First, we had several problems related to the power saving and screensaver features on the laptops we used for testing. Since we were transporting these computers with the lids closed and not directly interacting with the keyboard or mouse, we had to make sure these features were fully disabled. We neglected to do so in the first of our tests, which lead to lose of screen capture information. We also lost valuable testing information during two of the tests when the USB cables connecting the BlackBerry device and camera to the laptop became disconnected.

We also discovered that the software we used to capture the BlackBerry screen, BBScreenStream, is not entirely reliable. First, it often captures a white frame instead of the BlackBerry screen during periods of fast scrolling. We also soon realized that using BBScreenStream causes the BlackBerry device to perform slower than normal, especially when responding to user input. This problem is most severe when BBScreenStream is configured to capture the screen ever 0.5 second. Therefore, we worked around this limitation by ensuring that we captured at the lowest frame rate the software allowed, one frame every 2 seconds.

In addition to the indoor cellular communication problems we discovered, we also had occasional problems with the GPS reception of the BlackBerry device, most severely during our second pilot test. Initially we believed this was due to mounting the USB camera on the top of the device, perhaps interfering with antennae. However, later tests proved that this was not the cause of the problem, and we can only guess that it had something to do with the overcast weather on the day of the second pilot test. This disruption forced our test subject to complete the tasks in an entirely unique way, however, perhaps leading to richer result data overall.

Our technical problems are rounded out by the audio interference noise we recorded at all times that the BlackBerry device was communicating with the cellular data network. This noise presents its self as a high-pitched buzz that occasionally makes it difficult to hear what the subject is saying. While we only experienced this noise when recording audio using the laptop and headset-style microphone, we believe that the overall audio quality improvement achieved by using this method (vis-a-vis the external audio record and microphone) make it worth the annoyance.

Additionally, we also experienced several problems related to conducting our user tests in an outside environment. All of our tests were conducted during the cold Michigan spring, and our first pilot user test was actually performed in the rain. Luckily, we had an umbrella and extra hand available to ensure the comfort of our test subject. Furthermore, the noise from the wind and from passing buses and cars often made it difficult to record clear audio during our test. However, this problem was ameliorated by the use of the headset-style microphone.

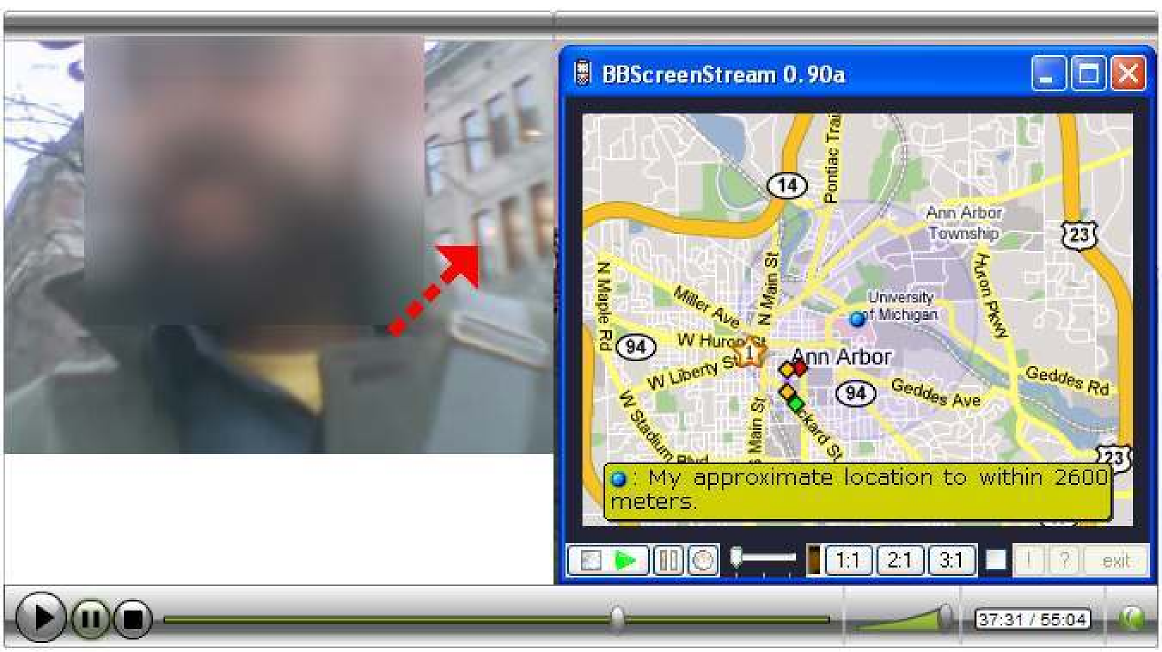


Figure 4: An outdoor test on an overcast day when camera could capture user's facial expressions (left). A snapshot of device captured by BBScreenStream (right).

Finally, our biggest data loss was due to the unexpected appearance of the Sun in the East Michigan sky. While we tested our USB camera both indoors and outdoors, we performed our primary outdoor testing during our second pilot test, which, as mentioned, occurred on a particularly overcast day. During our later user tests, the increased sunlight in the sky caused the image from the USB camera to be almost entirely washed out. This problem could have likely been corrected by purchasing a higher-quality digital USB camera equipped with a variable aperture and light sensor.

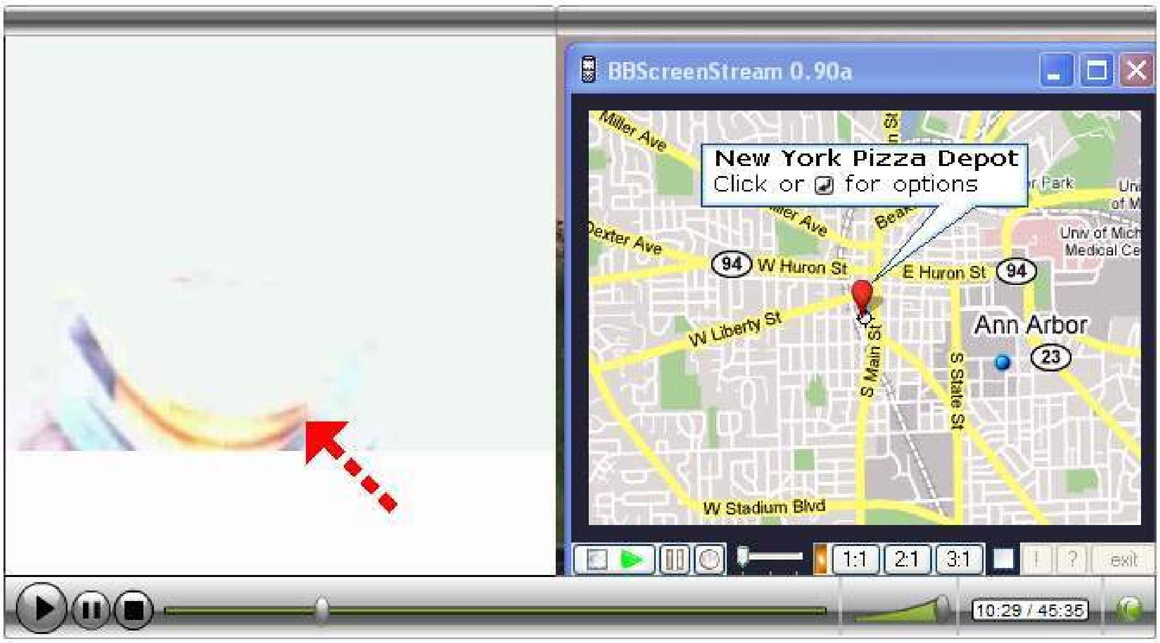


Figure 5: An outdoor test on a bright sunny day when camera failed to capture user's facial expressions (left). A snapshot of device captured by BBScreenStream (right).

4. TEST SUBJECT POPULATION

As outlined above, we conducted two pilot tests, and five actual tests (though considering out technical problems, the first two actual tests still were still somewhat provisional). All but one of our subjects was male, which is in keeping with the gender distribution of users that we found in our survey (69% male in the survey, 85% male in the user tests).

5. KEY FINDINGS

5.1 Problems Summary

As outlined in the Table 3 below, most users experienced similar challenges in performing the tasks described in Table 1. above. It is worth noting that although our users did encounter some problems that matched the results of earlier heuristic evaluation, these user tests primarily helped in discovering the problems that were not uncovered by any means before. All identified problems were ranked along Jacob Nielsen's rating scale for usability problems. Utilizing this standardized scale allowed for a convenient way of prioritizing our findings, and a consistent means of combining the results obtained from each subject.

Table 2: Priority Rating Scale

Priority	Definition
0	I don't agree that this is a usability problem at all.
1	Cosmetic problem only: need not be fixed unless extra time is available on project
2	Minor usability problem: fixing this should be given low priority.
3	Major usability problem: important to fix, so should be given high priority.
4	Usability catastrophe: imperative to fix this before product can be released.

The following table lists the seven primary issues identified via the user tests. Usability issues are ranked from most to least severe, and reference the broad categories usability, functionality, and aesthetics:

Table 3: Summary of Usability Problems

#	Issue	Priority	Category
1	Application does not orient to direction of travel	4	Functionality
2	Lack of audible directions	4	
3	Loading speed is inconsistent	3	
4	Application does not give feedback on successful or unsuccessful progression of route	3	Usability
5	Lack of contextualized numeric feedback – Issue 1	2	
6	Lack of contextualized numeric feedback – Issue 2	2	
7	Search box has poor affordance	2	Aesthetics

5.2 Details of specific problems

#	Issue	Priority	Category
1	Application does not orient to direction of travel	4	Functionality

Explanation: Route directions do not recognize users' moment-by-moment directional orientation, and start with standard instructions which sometimes sound ambiguous. For example, as can be seen in Figure 6, irrespective of user's momentary direction on S University Ave, user gets a standard direction stating 'Turn right at S State St - 0.2 mi.' It was observed that the first reaction of a typical user was to figure out what exactly was the correct 'right turn.' Users have to take a chance with common sense in taking the first step of directions. Further, even if users do succeed in the first step, ambiguity prevails. As can be seen in Figure 7, once a user selects to move 'right' on S University Ave, following direction on the map says 'Head west on S University Ave toward Tappan Ave - 0.3 mi.' One of the users' most notable think-aloud expression was 'Now how do I know which way is West?' Some users also desired to have 'compass points indicated on the map (so that they could know which way was west'.) It is evident that ambiguous directions introduce undesirable cognitive load as the application acts in a standard way rather altering the directions to best serve the user's context and momentary need.

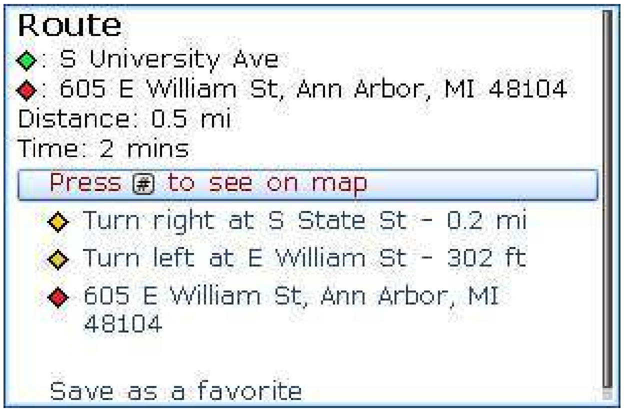


Figure 6: Irrespective of actual momentary position, user gets a standard direction to 'turn right.'

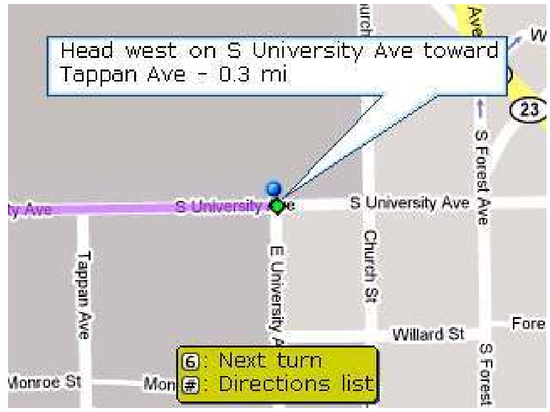


Figure 7: An initial direction to 'head west' confuses the user.

Possible Solutions: Deliver initial directions based on user's momentary directional orientation. Include flexible compass points on the map that change continually with the users' movement, giving users the real-time information and directions based on his current directional orientation.

#	Issue	Priority	Category
2	Lack of audible directions	4	Functionality

Explanation: Once the users find directions on the GMM, their primary goal is to reach for and recognize the destination in the physical environment. They use mapping tool as a supplementary artifact only to support their cognition about the physical environment. However, it was observed that GMM's hard interface, coupled with numerous directions, icons, status messages consumes all of users' limited cognition. When they could derive more concrete directional hints from the physical environment, they actually spent their limited cognitive capacities in wrestling with the interface. It was both observed and validated from think aloud expressions that users desperately wanted to 'hear' the directions aloud rather than reading them on the device [Several users frequently vocalized the desire to free themselves from having to look at screen.] Worse yet, extreme sunlight, or extreme stormy weather makes it all the more difficult to concentrate on the vaguely visible screen – the conditions that are surely a candidate for audible assistance than visual.

Proposed Solutions: Include audible directions with each search result. The audio directions should also be stored in the device cache for repeated hearing. Further, make the directions contextual with appropriate warnings, feedback, and information about successful or unsuccessful route traversing.

#	Issue	Priority	Category
3	Loading speed is inconsistent	3	Functionality

Explanation: Once users get accustomed to the application interface, their primary and most recurrent task is to search for, and pan around, the maps until they find the desired location. However, inconsistent loading speed was observed to disrupt the primary system functionality when one of the users tried 'the search task' indoor in overcast weather. This observation was also confirmed when the user reported to have keyed inputs at a predictive speed but only to receive unexpectedly slow response. Worse, the slow behavior is inconsistent, resulting in user frustration, input errors and greatly unexpected task completion time. The resulting disconnection between user's actions, expected results and actual results is a source of bad user experience.

Proposed Solutions: Make the application detect poor weather and signal, and make it visible to the user through warnings and status messages. Include specific, quantifiable, time-bound information along with the glass hour icon. E.g. "Trying to retrieve information in 5s..."

#	Issue	Priority	Category
4	Application does not give feedback on successful or unsuccessful progression of route	3	Usability

Explanation: At least three users wrestled with this issue in one or more of the following ways: Firstly, as evident from Figures 8, application does not issue explicit warning or feedback when users stray from the specified route. The feedback is rather implicit. The application expects the user to know that when the blue dot strays out of the route lines, user should infer that he is straying. For users already wrestling with a difficult interface, this obscuring feedback results in increased cognitive requirement, and raises the scope for uncertainty, delays and errors.

Secondly, there is no feedback about important milestones that the user just passed or about the success message when user reached the destination.

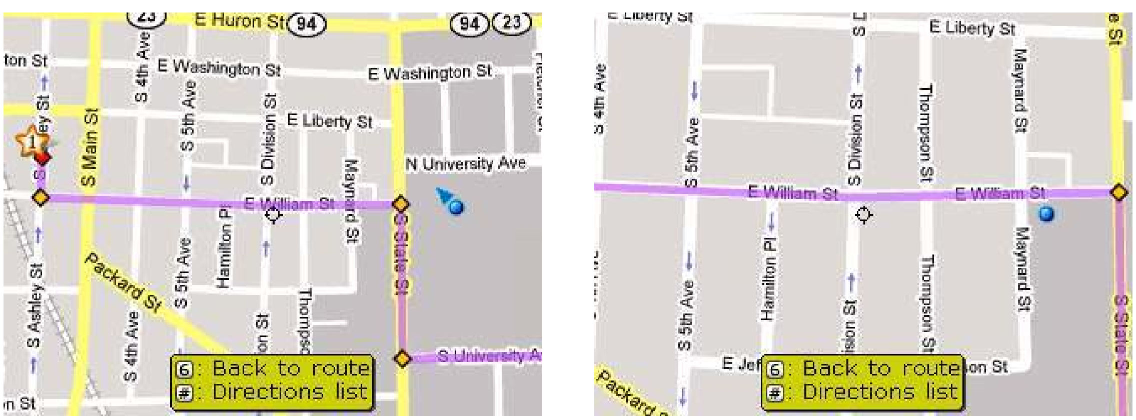


Figure 8: A blue dot indicating that user is straying from correct route (shown in purple lines.)

Possible solutions: Although hard to implement, an ideal solution will be to let the phone vibrate, along with relevant screen message, when user strays from the route or when he successfully reaches the destination. If tactical feedback is not possible, it is at least imperative to provide visual and textual feedback along each milestone of the route.

#	Issue	Priority	Category
5	Lack of contextualized numeric feedback – Issue 1	2	Usability

Explanation: The existing data transfer feedback only informs about how much data is loading. Unlike traditional web scenario, mobile users have not yet established semantic understanding of how much data is too much or too less. For example, as shown in Figure 9, knowing that 115KB worth of data is loading gives users no indication of how much time it is going to take. Further, it also does not inform users what percentage of 115KB of data has been loaded, and how much is remaining. This observation was validated by one of the user's think-aloud expression which said, "When transferring data, quantity being transferred is not given to me in context of total amount."

Figure 9: Data transfer feedback only informs how much data is loading. .

Proposed Solutions: Give explicit information if the data being loaded is heavier for given bandwidth of signal. Include more concrete message by displaying what percentage of data has been loaded, or how many seconds are remaining.

#	Issue	Priority	Category
6	Lack of contextualized numeric feedback – Issue 2	2	Usability

Explanation: It was observed that some users find it more valuable to know approximate arrival times to final destination and intermediate milestones en route, rather than just knowing how far the is the destination. "I always like to know what time I will reach there. Knowing that my destination is 0.3 miles away does not help me much"
Proposed Solutions: Include a constant feedback space to display how much more travel time is left. Alternatively, intermittent feedback pop-ups at each milestone could also keep the user informed. Current feedback message, displaying number of miles, can also be extended to include information about travel time.

#	Issue	Priority	Category
7	Search box has poor affordance	2	Aesthetics

Explanation: Our observation showed that new users showed some difficulty in understanding the affordance of search box. We witnessed two users who initially thought 'Enter a new search' tab shown in Figure 10 is the actual search box, but only figured with little trial-and-error that this tab is in fact a link to actual search box which is shown in Figure 11. These were reasonable assumptions to make on part of users as this tab gives an affordance of a text box seen in DHTML-supported modern-websites and/or in Apple computers.

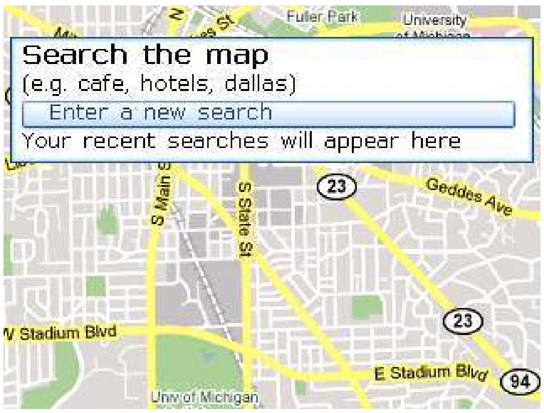


Figure 10: 'Enter a new search' looks like a text-box for entering search terms.

Figure 11: Actual text box is one screen ahead with unfamiliar affordance.

Possible Solutions: This is an issue of information aesthetics, and can be easily solved by simply changing the style in which the link to search box is presented. Various visual concepts can be tried and validated by interviewing/ observing with sizeable number of new users on how they respond to the proposed new style. But, importantly, whatever new style is chosen, it should either have its own unique affordance or should be similar to a link that indicates actual search action on the following screen.

5.3 Some positive surprises

Our evaluation also identified some features of GMM that users are apparently comfortable with. Some of them are as follows:

- Contrary to our own observations, our users could easily discover the blue arrow indicating their direction.
- Users could learn key commands quickly (4 for backward directions and 6 for forward directions, and # for favorites; and i and o for zoom-in and –out, and # for favorites)
- Users were not bothered by the zoom instructions screen when they pressed zoom. Two of our users learned it pretty quickly.
- Users appreciated that the application returns to the last state the user quit
- Many users also deliberately tested the applications ability to deal with vague searches, and several of them expressed pleasure when the application was able to find what they were searching for with limited input
- A few users also liked the color differentiation of route on map.

6. SUMMARY

This evaluation identified and prioritized seven problem issues. In order of severity, these are as follows:

1. Application does not orient to direction of travel
2. Lack of audible directions
3. Loading speed is inconsistent
4. Application does not give feedback on successful or unsuccessful progression of route
5. Lack of contextualized numeric feedback – data transfer issue
6. Lack of contextualized numeric feedback – arrival time issue
7. Search box has poor affordance

These problems, when addressed by developers, will greatly improve the functionality and usability of GMM. This study also assisted us in developing deeper and broader understanding of the mobile mapping applications in general. In addition to discovering new problems, these user tests also helped in confirm and clarifying some of the findings of previous heuristic evaluations completed by our team. If applied meticulously, this understanding will not only assist the development process, but also help in crafting more marketable, user-centered mobile mapping applications in future.

7. REFERENCES

Rudy Schusteritsch, Carolyn Y. Wei, & Mark LaRosa (2007). Towards the Perfect Infrastructure for Usability Testing on Mobile Devices. *CHI*, pp. 1839-1844

Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, & James A. Landay (2007). MyExperience: A System for In situ Tracing and Capturing of User Feedback on Mobile Phones, *MobiSys'07*, pp. 57-70.

Minna Isomursu, Kari Kuutti, & Soili Väinämö (2004). Experience Clip: Method for User Participation and Evaluation of Mobile Concepts. *Proceedings Participatory Design Conference,* pp. 83-92.

Eeva Kangas, & Timo Kinnunen (2005). Applying User-Centered Design to Mobile Application Development. *Communications of the ACM,* 48(7), pp. 55-59.

Nielsen, J. (1993). *Usability Engineering*. Academic Press, San Diego, CA.

Usability Testing. Retrieved March 2008 from http://en.wikipedia.org/wiki/Usability_testing

Nielsen J. (2000). Why You Only Need to Test With 5 Users. Retrieved March 2008 from http://www.useit.com/alertbox/20000319.html

8. APPENDIX

8.1 Consent Document

Introduction
Hello my name is ___________________________. I am a student in the School of Information course 622: Evaluation of Systems and Services. Thank you very much for coming. Please allow me to briefly explain why we've asked you to come in today and what exactly you will be doing.

You will be helping our team in evaluating how easy it is to use Google Maps for Mobile (GMM) on Blackberry 8310 for specific tasks. I will be asking you to perform some typical tasks with Google Maps for Mobile. Please complete the tasks as you normally would, for example, take the same amount of time and attention to detail that you would if you were performing the tasks on your own. Do your best but don't be too concerned about your results. Please be assured that we are interested in evaluating the behavior of the product, not you.

When you are completing tasks please talk aloud through the tasks. Tell me what you are doing and what you are thinking and feeling as you perform the task. For example, if my task is to open a file in Microsoft Word I would say out loud "First I go to File menu. I click File and scroll down to Open. After I click on Open a window pops up..."
Since our purpose is to simulate how someone would use the program on their own with no external help, I will not be able to answer any questions you have about the program or help you through any of the tasks.

Equipment
Let me demonstrate the equipment. For an indoor test, this computer has Blackberry and Google Maps for Mobile stimulators loaded on it. We are using the software called ‘Camtasia’ to capture screenshots of the application, and the movement of mouse on the phone simulator screen. We have another camera that will be recording throughout the test. For outdoor tests, the Blackberry device will be connected to our computer through a USB. We will be using the software called ‘BB Stream Shooter’ to capture your actions on the device.

These recording will never be made public and will only be utilized by our team members as a reference for analysis. We will also be capturing audio data.

We ask that you refrain from moving out of the test simulating window. If it is necessary that you browse out of the window to complete the task, please feel do so, but clearly explain why.

Our equipment in-charge may need to interrupt the task in case any problems are noticed.

We have no affiliation with either Blackberry or Google, who are the makers of these products.

Please let me know, if at any time you would like to stop this session. You will still be compensated if you decide to quit in the middle of the session.

Do you have any questions?
If not, let's begin by having you sign this consent form.

Consent Form
We will be recording video and audio of your session. These videos will never be made public and will only be utilized by our team members for GMM analysis.

Please read the statements below and sign where indicated.

Thank you.

I understand that videotape and audiotape recordings will be made of my session. I grant the School of Information 622 Evaluation Team to use these recordings for the purposes mentioned above, and waive my right to review and inspect the tapes.

Print Name: ____

Signature: ____
Date: ____

8.2 Usability Testing Protocol

1. Setup Tasks

Equipment wizard:

1. Set up Camtasia to record face
2. Set up Blackberry (connect to machine, start BB Screenstream, open blackberry to GMM, reset GMM, accept liscence agreement)
3. Open a browser window to the Pre-questionnaire (http://s-tp316-38637.sgizmo.com)
4. Set up audio recording equipment for outdoor test
5. Set up webcam

Note taker:
1. Get out note taking forms
2. Prepare a timer (on your computer or otherwise)

Moderator:
1. Prepare consent forms
2. Place question list next to computer**

2. When Participant Arrives:

Moderator:

1. Read orientation to the test

2. Give them consent form
3. Ask them to take the pre-questionnaire
4. Give orientation to the device
    1. blackberry
    2. roller ball (scroll and click)
    3. back button
    4. return button

3. Start Indoor Task:

Equipment wizard:

1. Close pre-questionnaire window
2. Starts BBScreenstream
3. Start Camtasia
4. Start kitchen timer**

Note taker:

1. When camtasia starts running, start your timer and start note taking

Moderator:

1. Guide participant through tasks, for warm-up task, give feedback about device

<u>Warm up task:</u>
Search for 635 Fifth St., in Ann Arbor, MI.

<u>Actual task:</u>

1.
 a. You and your friends are planning on going out tonight for dinner at the New York Pizza Depot on East William St. You've never been, so you decide to look it up on your phone. Tell us when you've found it.
b. Now save it as a favorite in Google Maps for Mobile to use for later. Please let us know when you've succeeded in this task.

2. You know you'll need to find parking when you go, so you decide to look around the area of the restaurant to see what might be nearby. Pan around to find what looks like the nearest parking garage. Please let us know when you've succeeded in this task.

4.5 Debrief Questions

Moderator

1. Administer indoor task post interview

    1. What was your general impression of the actual task (not the warm-up task)?
    2. Did you encounter any problems searching for the Pizza Depot?
    3. Did you encounter any problems saving it as a favorite?
    4. Were there any features of the application that you think could be improved?
    5. Were there any features of the application that you found particularly useful?

4. Prepare for Outdoor Task

Equipment Wizard:

1. Stop Camtasia
2. Keep BBScreenshooter open, start recording
3. Turn on audio

5. Administer Outdoor Task

All: Walk downstairs together

Wizard:

1. Set up user in the equipment

Moderator:

<u>Actual task:</u>
Using your current location, find a route to the New York Pizza Depot, which you saved as a favorite, inside. Follow it there, noting each turn aloud on the route as you make it. Let us know when you arrive.

All: Walk back to library

6. Debrief:

Moderator:

1. Administer outdoor post interview

    1. What was your general impression of the task?
    2. Did you encounter any problems in creating a route from your current location?
    3. Did you encounter any problems in following a route?
    4. Were there any features of the application that you think could be improved?
    5. Were there any features of the interface that you found particularly useful?

7. Take down:

Wizard:

1. Export video from Camtasia
2. Save audio and sync with screen recording
3. Upload to Bart, send link to group with file name "622 Test #X"

Note taker:

1. Digitize your notes and email to person who did the tech

Data person:

1. Review digital files, take notes, send to tech person to upload to Bart

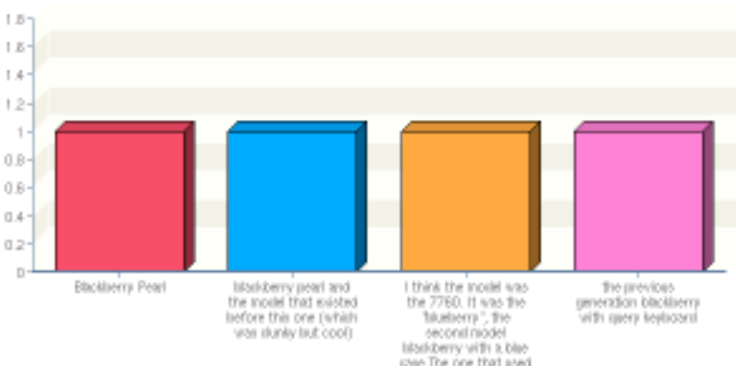
8.3 User Test Survey Responses

![SurveyGizmo — online surveys & data collection — www.SurveyGizmo.]

Report: Response Summary Report

Survey: User Test Pre-questionairre
Compiled: 03/27/2008

1. Which Blackberry device(s) do you have experience with? (please list)

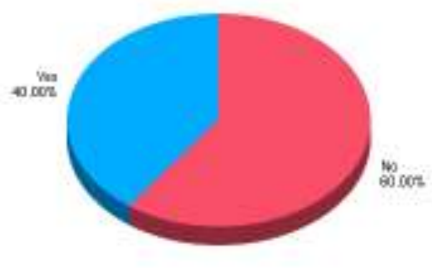


STATISTICS	
Choices Selected:	5
Total Responses:	5

SUMMARY

VALUE	COUNT	PERCENT %
Blackberry Pearl	2	40.00%
blackberry pearl and the model that existed before this one (which was clunky but cool)	1	20.00%
I think the model was the 7760. It was the "blueberry", the second model blackberry with a blue case.The one that used had a color screen, the first version had a black and white screen.	1	20.00%
the previous generation blackberry with query keyboard	1	20.00%

2. Have you ever used Google Maps for Mobile on the Blackberry?

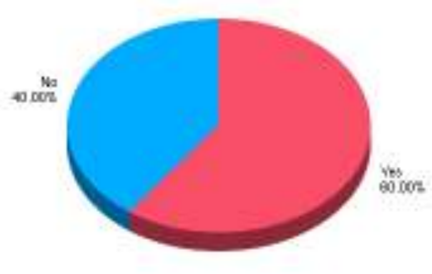


STATISTICS	
Choices Selected:	6
Total Responses:	6

SUMMARY

VALUE	COUNT	PERCENT %
No	4	66.67%
Yes	2	33.33%

3. Have you every used Google Maps for Mobile on another mobile device?



STATISTICS	
Choices Selected:	6
Total Responses:	6

SUMMARY		
VALUE	COUNT	PERCENT %
No	3	50.00%
Yes	3	50.00%

4. Do you use Google Maps on your personal computer?

STATISTICS	
Choices Selected:	6
Total Responses:	6

SUMMARY

VALUE	COUNT	PERCENT %
Yes	6	100.00%

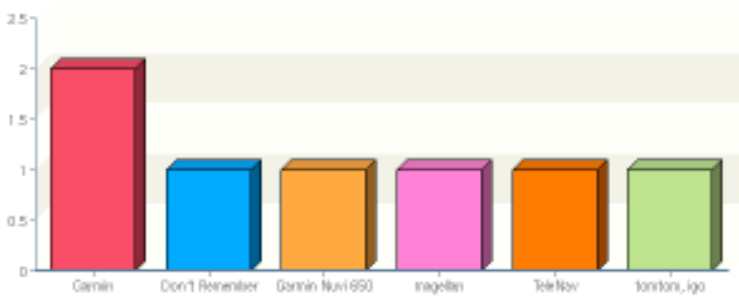
5. If you use google maps for mobile, please rate how hard or easy you find the following tasks: ()

ITEM	very easy	somewhat easy	neutral	somewhat difficult	very difficult	n/a	Total
searching for an address	20.0% 1	40.0% 2	20.0% 1	20.0% 1	-	-	5
searching for a vague location (e.g. business type or cross streets)	20.0% 1	-	60.0% 3	20.0% 1	-	-	5
creating a route	20.0% 1	80.0% 4	-	-	-	-	5
following a route	20.0% 1	20.0% 1	-	60.0% 3	-	-	5
Average %:	**20.0%**	**35.0%**	**20.0%**	**25.0%**	**0.0%**	**0.0%**	

Total Responses: 5

6. Do you have experience using mobile mapping tools other than Google Maps for Mobile?

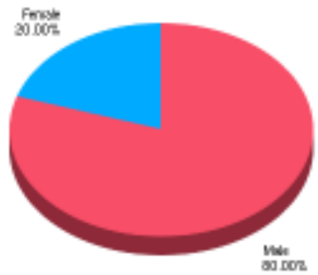
(If you do not have experience with other mobile mapping tools, please skip to the next page of the survey by clicking the Next Page button at the bottom of this page)



STATISTICS	
Choices Selected:	5
Total Responses:	5

SUMMARY		
VALUE	COUNT	PERCENT %
Garmin	2	40.00%
Don't Remember	1	20.00%
Garmin Nuvi 650	1	20.00%
magellan	1	20.00%
TeleNav	1	20.00%
tomtom, igo	1	20.00%
VZ Navigator	1	20.00%

7. Please indicate your gender:

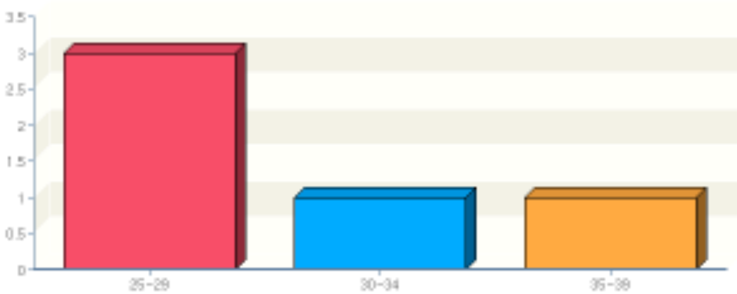


STATISTICS	
Choices Selected:	6
Total Responses:	6

SUMMARY

VALUE	COUNT	PERCENT %
Male	5	83.33%
Female	1	16.67%

8. Please select your age range:



STATISTICS	
Choices Selected:	6
Total Responses:	6

SUMMARY

VALUE	COUNT	PERCENT %
25-29	4	66.67%
30-34	1	16.67%
35-39	1	16.67%

Appendix: 9.

Please explain your experience with the device(s) listed above (how long you used it, how frequently you used it).

DATA	
CODE	VALUE
6364921	I used the device for 8 months, and made use of all of its planner features, Google Maps, frequent use of the web browser for news.
6376122	Was issued at my place of work. UCSD school of medicine, national center for microscopy and imaging research. Used it about for about 2 years almost constantly. Our group had about 12 of them.
6411002	The BB pearl that I use now is much improved over the previous. It is faster, easier to navigate (hated the side scroll from prior model). My job gave me the phone for email, web use, IM, and phone (of course). I've carried a BB for about 3 years and this newer model for about 1 year.
6444663	daily, for about 2 years, using email and phone services
6521110	Only occasionally (less than 10 times)

Appendix: 10.

If you answered yes to the previous question, please list which devices you have used Google Maps for Mobile on:

DATA	
CODE	VALUE
6364921	Samsung BlackJack II
6376122	Yes and no. It's google maps on the iphone, which is technically not google maps for mobile. But it is a mapping application by google for a phone.
6444663	windows mobile devices like HTC and palm treo