Pilot Usability Study

Project Title: ETA – Electronic Travel Assistance for Group 4.
Online Doc.: http://inst.eecs.berkeley.edu/~swlee/assignment4.html

Introduction

Electronic Travel Assistance (ETA) is a program designed to help travelers quickly access powerful features such as accommodations, dining, entertainment, maps, and other information, while at the same time, help to create a custom itinerary for any desired location. The goal of the project is to improve the overall experience of budget travelers by providing up-to-date information not found in tour books and the ability to manage those resources at hand. The program runs on Java Chai on Microsoft’s PocketPC handhelds.

We are now at the stage of evaluating the high-fidelity prototype that we have produced during our last assignment. The purpose of this study is twofold: first, to perform a usability test on the modifications to our interface since our low-fidelity experiment, and second, to use this gather information to enhance our current working model of the program.

Method

Our Participants were the following:

First participant was 20 yr., Mechanical Engineering major, 3rd year at Berkeley.
   Gender:     Male
   From:       Thailand
   Experience: No prior experience.

Second participant was 25 yr., IEOR major, graduate student.
   Gender:     Male
   From:       France
   Experience: Some experience with Sony Clie.

Third participant was 24 yr., Business major, 4th yr.
   Gender:     Female
   From:       United States
   Experience: Has used a Jornada before.

The equipment that we used in the study consisted of: a HP Jornada, JVC Digital Camcorder, paper, and pencil. We performed the study in a quite room in the International House, which offered us a relatively distraction-free environment. Jun greeted the subjects and introduced them to the study by giving them the consent and information forms. Bikas introduced them to the system by showing them a short demo of ETA. Fahd made observations, and Sean filmed the entire study on video tape.

Our Tasks were the following:

1. The first task assigned to the participants is to set up their profile. The goal of this task is to identify not only how easy the profile section is to navigate, but also to evaluate if the questions
asked reflect characteristics of the user that will be relevant and helpful when making recommendations. We insisted on avoiding asking questions that sound restrictive, and tried rather to get input that helps customize the recommendations without limiting the options of the user.

2. The second task is to find an accommodation, the user navigates through the accommodation section and uses the suggestions to locate a convenient accommodation. This section will be used frequently, so we were concerned about the ease of use and the completeness of the format for displaying information about the place. It includes price, address, and a review as well as options such as mapping the accommodation. We also were looking at how flexible and convenient the navigation between the suggestions, the details of a specific accommodation such as it map, and then back to the listing of accommodations.

3. The third task consisted of mapping a location. We think it is one of the most important features that will be used on a daily basis. We looked at how easy it was for users to input addresses using the keyboard interface of the Jornada. We also tried to replicate the standard way of displaying maps with zooming buttons, to be consistent with the standards the users might be used to.

Our Procedures were the following:

We followed a standard procedure with all participants. First, one member of the group was assigned the task of greeting the user and helping him/her go through the release forms. A second member of the group explained the general dynamics of the user interface and the purpose of our tool as a travel assistant. The introduction was meant to give the participant a general feel for the interface without revealing the way tasks are accomplished. The user was also instructed on the procedure used during testing, especially the absence of assistance with tasks and the importance of speaking loudly as the task execution progresses.

A third member of the group provided the tasks for the participant; he only interfered when a task was finished and to provide the next one. Throughout the testing process, a fourth member videotaped the complete interaction of the user with our interface; this allowed us to review the performance of participants with greater details and without loss of information after all the participants were done. The taping of the three tests allowed us to save valuable input from the users in the form of critical input. Once participants finished the third task, they were thanked for their help. We also insisted on getting more input from the participants by asking about their general impressions about the feel and look of the interface. We also asked for suggestions to improve on the interface and its features.

**Test Measures**

We measured the following:

1. The time necessary for accomplishing the tasks: we believe that comparing times required by the participants to accomplish a certain task can reveal anomalies and inherent complications if that task takes longer than our expectations. We kept in mind, however, that some users were not used to the keyboard interface and required more time to input data while others tended to talk more during the tasks.
2. The number of errors: We kept track of the number of times users went down the wrong track instead of the succession of steps we expected. A high number of errors indicates that the interface for that task is not intuitive and requires some changes.

3. The Critical input: Using the tape of the performances, we were able to analyze subtle indications of satisfaction or discomfort displayed by the users through comments, body language, we believe that will allow us to address the issues of any part of the user interface that may cause a high level of discomfort to the user.

Results

<table>
<thead>
<tr>
<th>Participant</th>
<th>Task #</th>
<th>Time</th>
<th># of Errors</th>
<th>Critical Incident ( severity rating / implementation difficulty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3 min 40 sec</td>
<td>1</td>
<td>a. Tester exited from our program after the second screen. (3 / -)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Tester wasn’t clear whether “price” was the price for the whole trip or just for a day (2/0)</td>
</tr>
<tr>
<td>2</td>
<td>0:42</td>
<td>0</td>
<td>0</td>
<td>a. Tester said, “How do I pick a hotel?” → failed to recognize the detail hyperlink at first (1 / 1)</td>
</tr>
<tr>
<td>3</td>
<td>0:57</td>
<td>0</td>
<td>0</td>
<td>a. Tester successfully performed the function in a different way than what we thought → went back to the screen from task 2 in order to find a map instead of going back home and clicking on the map icon (0/0)</td>
</tr>
<tr>
<td>1</td>
<td>4:10</td>
<td>1</td>
<td></td>
<td>a. Tester failed to see that screen was half way scrolled down (side effect from previous test) and didn’t fill out some fields. (3 / 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Complained that he can’t see what he was typing in the textbox since the Jornada’s keyboard was covering the textbox. (3 / 4)</td>
</tr>
<tr>
<td>2</td>
<td>1:14</td>
<td>0</td>
<td></td>
<td>a. Tester said “Cool, I can see the map of it”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Mentioned that it would be nice to have an option of making a reservation wirelessly if they had a wireless connection.</td>
</tr>
<tr>
<td>3</td>
<td>2:5</td>
<td>0</td>
<td></td>
<td>a. Tester took some time to figure out how to change Jornada’s system keyboard from character mode to numerical mode. (2/4)</td>
</tr>
<tr>
<td>1</td>
<td>2:27</td>
<td>2</td>
<td></td>
<td>a. Instead of clicking on NEXT button, she clicked on Tools menu before figuring out that NEXT button was accessible by scrolling down. (1/2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b. Failed to see that the screen was scrolled down when she first clicked on it. Didn’t fill in some of the fields herself; It was already filled out from previous test. (3 / 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c. Failed to recognize that more than one selection could be picked. (2/1)</td>
</tr>
<tr>
<td>2</td>
<td>0:37</td>
<td>0</td>
<td></td>
<td>a. Tester didn’t fill in some of the fields.</td>
</tr>
<tr>
<td>3</td>
<td>1:15</td>
<td>0</td>
<td></td>
<td>a. Tester didn’t fill in some of the fields.</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>3:25</td>
<td>1.3 Err</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0:51</td>
<td>0 Err</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1:25</td>
<td>0 Err</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Notes on bottom-line data:
  - Testers took most of their time filling in the textboxes using Jornada’s on-screen keyboard. Also, it took them longer to complete a task if they were more verbose about what they were thinking. So, the time to complete a task might not correctly interpret the how much thinking they needed to put in to perform the task, and how difficult it is to use the particular feature.
  - Tester 1 had never used Palm or Windows CE devices. Tester2 had some experience with Palm devices (Sony Clie) but not with the Windows CE devices, and tester 3 used to own HP Jornada.
# of Errors included any minor button click that was not necessary as well as failing to recognize/read/fill-in parts that was pre-filled due to the reason below.

During the pilot tests, we used the same Jornada and when users clicked on some pages, the screens were scrolled down already, thus confusing users to recognize the information on the upper part. In real life situation, it is not very common to have two different users using one Jornada.

Severity rating and implementation difficulty on the critical incidents range from 0 to 4.

Process Data

- We noticed that the results from the pilot test were enormously better than that from the contextual inquiry. For example, the simplified tool bar and use of icons on the home screen dramatically cut users’ reaction / process time to perform tasks. However, the critical incidents listed on the table above also revealed some designs that could be further polished particularly for novice Windows CE users. Most importantly, we recognized that we needed to provide more rigorous error checking schemes and confirmation screens. Also, we needed to integrate the application with Windows CE better:

Exiting the Application:

- The most severe error was from the first tester who exited the application in the midst of creating his user profile. It would have been helpful to him if the application prompted the user with warning popup screen asking, “You are existing ETA. Are you sure you want to continue?” with Yes and No buttons in addition to checkbox reading, “Don’t prompt me with this message again”.

- Similarly, it would be helpful for novice users to be able to give them overview of the function when they click on a button such as profile. The application would show them an extra introduction screen explaining what they can do using the particular button or icon. For example, if the user clicked on a profile button, the next screen would tell the user what he/she could do in profile, why creating his profile is useful, as well as the estimated process/time it will take them to create a profile. The user would have the checkbox that would let them disable the explanation screen once they know what it does.

Integration with Jornada’s Keyboard:

- Two testers out of three complained that, when Jornada’s keyboard is clicked, it covers half of the screen; sometimes covering the textboxes that users were entering information on. In order to fix this problem, we need to scroll the active textbox to the top of the screen in order for the users to see what they are typing using the on-screen keyboard. In addition, we could implement (if Windows CE provided SDE tools for it) a function so that the keyboard will be set to appropriate mode: numerical or character mode depending on the textbox type that user is using. For example, if the textbox required the users to enter the daily budget in dollars, keyboard would popup on numerical mode.
Error Checking:

- Some of the users failed to enter some of the data because of scrolled down pages. It would be more easily recognizable for users if the screen scrolls up to the top part every time the screen is clicked whether or not it was scrolled down to the bottom when last used. In addition, the application should check if all of the required fields were entered along with the confirmation screen after entering all the data.

Better On-Screen Instruction:

- One tester clicked on the tools button to proceed to the next page shortly before clicking on the NEXT button. To avoid these scenarios, the screen should more explicitly explain how to proceed to next screen.

- In addition, some of the testers failed to recognize that more than one checkbox could have been picked. For example, the application asks the user what his/her music types are, and it has choices with checkboxes such as classical, jazz etc. We actually had this problem with our first design and we changed the selection menu to checkboxes. However, it is still not clear enough for some users to be able recognize that multiple selections could be made. It would be more instructional if we told them explicitly in words, “You can choose more than one selection (or none).”

- Also, a tester didn’t know how to get more information about a hotel. It would be helpful to tell users they could get more information by clicking on “detail” hyperlink.

Discussion

The results of the hi-fi pilot test revealed a remarkable improvement over our initial lo-fi prototype. Perhaps the most important was a considerable improvement navigating through the interface. In our lo-fi prototype, navigation was done through a pull down menu, with many of the sections hidden in submenus. This created a startling problem for users during our lo-fi testing. The menus were redone in the hi-fi prototype to make it more clear how to reach a section, and a new main page was added to the interface with icons that provide clear links to each of the sections of the interface. The result, as seen in the pilot test, was that all of the participants were able to clearly find the appropriate section of the interface to complete the necessary task. Furthermore, a ‘home’ icon was placed on each page of the interface, thus allowing users to quickly return to the main screen. The results indicating the success of this design were evident in that none of the users need assistance beginning a task, and that all of the tasks were completed successfully on the first attempt, with one exception (discussed later).

Another positive result from the pilot test was the combining of the mapping feature of the interface with the mapping features of the listing sections. Users can pull up a map of a particular listing (hotel, point of interest, restaurant …), as well as a get a map given an address. The success of having a single map section (with links from two places) was evident in the results with participant 1 completing task 3 by using what he remembered from task two, that he was able to link to the map section from the accommodations section.
Our most significant design flaws were found in the profile section, with the most severe being that participant 1 exited from the application by mistake. The most likely reason explaining the results for the first task are that the users were not able to clearly see what they were doing. This was due to scroll bars and the Jornada keyboard hiding areas of the screen the user needed to see when typing. The most simple fix, first of all, is to move text entry fields higher up on the screen so that when the Jornada keyboard is activated, the text fields are not covered by the keyboard. Second of all, eliminating scrolling where ever it is at all possible will eliminate the problems the users had because they could not see all the on screen information, such as page titles or ‘next’ and ‘previous’ buttons. Participant 1 exited the program because he was not sure if he was in the right section, even though he clicked the correct icon from the main page. This was because the page was partly scrolled down covering the title of the page. Other participants did not fill in all of the information asked of them because they did not see to scroll down.

Finally, our pilot testing continued to reveal minor problems as a result of unclear, or incomplete directions or information. For example, none of the participants knew that they could choose multiple selections in the profile task for check boxes. Refining the on screen instructions and text was something that we spent a lot of time on from our lo-fi prototype to our hi-fi prototype. However, continued refinements are necessary to provide dialog that is both clear and concise.

To conclude, the pilot testing revealed a dramatic improvement between our lo-fi and our hi-fi testing. Many of the major interface decisions proved to provide a much more usable interface. However, the numerous minor flaws still yield an interface needs to be refined for our next iteration of design and testing.