Geopod Project

Usability Test Plan

Blaise W. Liffick, Ph.D.

November 23, 2010

1 Table of Contents

2. Document Overview

This document describes a test plan for conducting a usability test during the development of the Geopod software system. The goals of usability testing include establishing a baseline of user performance, establishing and validating user performance measures, and identifying potential design concerns to be addressed in order to improve the efficiency, productivity, and end-user satisfaction.

The usability test objectives are:

1. To determine design inconsistencies and usability problem areas within the user interface and content areas. Potential sources of error may include:
    a. Navigation errors – failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow.
    b. Presentation errors – failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities.
    c. Control usage problems – improper toolbar or entry field usage.
2. Exercise the application under controlled test conditions with representative users. Data will be used to assess whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.
3. Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

Users for this application are earth science students, primarily at or above the sophomore level. Students from this group will be selected to participate in a timed usability study. A discount usability[i] approach will be used with 14 test subjects, seven of whom are female; all of these participants use a mouse with their right hand. Additionally, two left-handed mouse users will be given the same set of tasks as in the 14 timed sessions, but will instead be given a "think aloud" process to follow rather than timed.

Testing will take place in the Adaptive Computing Lab of the Department of Computer Science, Millersville University.

3. Executive Summary

Geopod Functionality Outline:

1) Navigation
    a. Use keyboard and mouse
    b. Autopilot
    c. Compass
2) Noted locations
    a. Note location
    b. View and edit noted locations
    c. Load/save file with locations
3) Dropsonde
    a. Add/remove

    b. Use history
4) Particle imager
    a. Deploy
    b. Cycle images for current category
5) Grid points
    a. Change density
    b. Toggle on/off
6) Parameters
    a. Primary display
    b. Overflow
    c. Drag and drop
    d. Parameter addition/removal
7) Geocoding
    a. Forward (Location to lat/lon/alt)
    b. Reverse (lat/lon/alt to location)
8) Help
    a. View key commands
    b. View mouse commands

It is worth emphasizing that this is not intended as a validation test of the correct behavior of the Geopod controls, nor a test of the user's abilities, but is an evaluation of the effectiveness of the Geopod interface from the user's perspective. Usability goals include that the user will be able, after minimal training, to perform a series of guided activities within a "reasonable" amount of time, with a minimal number of errors, instances of "dead ends," and resorting to the help menu. The reasonableness metric will be judged against the need for students to be able to accomplish a certain amount of work using the Geopod system during a typical lab period. A "control" study was done by having an expert user perform the timed trials as well, to provide a lower bound on the times to complete the tasks.

Each test subject will be given four scenario test trials to complete, with a video snapshot of each trial taken. These snapshots will then be analyzed to compute time for task completion, error rates, instances of dead ends, and number of times the user activated the help menu.

4. Methodology

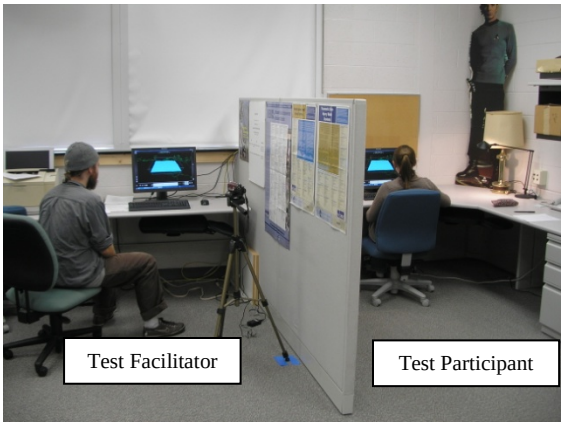
Two types of usability testing are being employed for this project. The first is an expert heuristic evaluation. In this case, an expert in human-computer interaction (Dr. Blaise Liffick) will perform a system walkthrough of the system, looking for potential problems based on his extensive experience in HCI.

The results from the expert walkthrough will be used to inform some elements of the usability study, as well as to provide general recommendations for improvements to the system.

The second type of testing is a standard usability study using a small group of test participants selected from the population of potential users of the system.

There will be 14 primary test participants, all of whom will be asked to complete the same set of four test trials, with half of the participants of each gender. The order of the trials will be the same for each participant, as affects of learning bias are not a concern in this study; indeed, it is assumed that participants will learn something from each trial that could prove useful in subsequent trials. Each test participant will be asked for basic demographic information including age, handedness, gender, and experience with previous 3D navigational systems. They will also be given a satisfaction assessment as a post-test to gauge their level of satisfaction with the interface.

The test system is housed in a cubical within the Adaptive Computing Lab of the Department of Computer Science, Millersville University. It is enclosed on three sides. One of the walls of this cubical is a 5' high partition. The test subject is positioned at a desk within the cubical. The test facilitator will sit outside the cubical on the other side of the partition, unseen by the test subject. The facilitator has a monitor that is a mirror of the test participant's monitor, as well as an active keyboard and mouse with which to interact with the system during trial setup and end-of-test housekeeping tasks. This also allows the facilitator to observe and record the action of the test without in any way being in direct contact with the test subject, either verbally or visually.



Recording will be done with both a video camera trained on the mirrored monitor and through screen capture software (using CamStudio), and through keystroke logging (provided by the Geopod system itself). The redundant recording mechanism is being used to ensure that a valid record of each trial is actually recorded. Note that the video capture mechanism must ensure that there is no performance impact on the Geopod system. Furthermore, the Geopod system is equipped with a logging mechanism that records time-stamped (accurate to at least the $1/10^{th}$ second) events detected by the system, including all keystrokes, button presses, etc.

Timings and error counting will be done after the tests through protocol analysis of the captured video, ensuring that such measurements are consistently taken. This will be accomplished by using video software (QuickTime) that allows video clips to be viewed accurately with a time indicator, measured to at least 1 second; the software allows frame by frame advancement of the video. In any instances where the video record does not adequately indicate tasks initiated by the test subject, the Geopod log file will be consulted to verify the initiation and/or completion of the task.

A facilitator will conduct each test. Acting as facilitators will be students from Dr. Liffick's CSCI 425 (Human-Computer Interaction) course. They will be responsible for initial setup of test conditions prior to each trial, managing the video recording mechanisms, note taking during the trials, and end-of-test housekeeping tasks. Dr. Liffick will observe every test to ensure that the student facilitators follow the test protocols developed by Dr. Liffick, in order to provide consistent conditions for all of the tests.

4.1 Participants

The 14 test participants will be recruited from current earth science courses. They include a representative number of earth science students from the sophomore, junior, and senior levels. There are an equal number of male and female test participants. They are expected to be sufficiently versed in meteorology principles such that the terminology and tasks used in the Geopod usability tests will be familiar to them – their knowledge of earth science is not being tested, however, only their ability to use the Geopod system effectively to complete the tasks.

The participants' responsibilities will be to attempt completion of a set of representative task scenarios presented to them in as efficient and timely a manner as possible, and to provide satisfaction feedback regarding the usability and acceptability of the user interface. The participants will be directed to provide honest opinions regarding the usability of the application, and to participate in post-session subjective questionnaires. Test participants will each be compensated $25 for their time.

4.2 Training

The participants will receive an overview of the usability test procedure, equipment and software and given approximately 30 minutes of hands-on experience with the system. They will receive initial training on the use of Geopod functions within one week prior to the usability study, and will be allowed 10 minutes of "exploration time" at the beginning of the usability test to refamiliarize themselves with the controls.

4.3 Procedure

Participants will take part in the usability test at the Adaptive Computing Lab (Roddy Hall room 140) in the Department of Computer Science of Millersville University of Pennsylvania. A computer with the Geopod application and supporting software will be used in a typical lab environment, except that the system is housed within a cubicle for privacy. The participant's interaction with the application will be monitored by the facilitators seated in the same lab, but on the other side of a 5' partition. The sessions will be observed via the use of a mirrored monitor, connected directly to the test computer. The test sessions will be videoed, the display will be video captured, and user keystrokes and other Geopod events will be logged and time stamped.

The facilitator will follow a General Testing Protocol (Appendix I) that describes the overall testing procedures. They will brief the participants on the Geopod application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant (see Appendix II – Test Facilitator Protocol). Participants will

sign an informed consent (Appendix III) that acknowledges their participation is voluntary, that participation can cease at any time, and that the session will be videoed but their privacy of identification will be safeguarded. The facilitator will ask the participant if they have any questions.

Participants will complete a pretest demographic and background information questionnaire (Appendix IV). The facilitator will explain that the amount of time taken to complete the test tasks will be measured and that any additional exploratory behavior outside the task flow should not occur until after task completion. At the start of each task, the participant will read the task description from the printed copy and begin the task when signaled to do so. Time-on-task measurement begins when the participant starts the task by maximizing the Geopod application from the system task tray. The facilitator follows the Trial Protocol (Appendix V) for each of the four test trials.

After all task scenarios are attempted, the participant will complete the post-test satisfaction questionnaire (Appendix VI). The facilitator continues following the Trial and General protocols to complete housekeeping tasks related to storing all electronic files appropriately, reinitialize the Geopod system, and continue with the next trial.

5. Roles

The roles involved in a usability test are as follows:

Trainer
Provides training overview prior to usability testing.

Test Facilitator

- Provides overview of study to participants
- Defines usability and purpose of usability testing to participants
- Conducts the test trials
- Records any useful information not readily captured on any of the video recordings or the automated data log

Faculty Observer

- Ensures that all test trials are conducted according to the stated protocols, in order to eliminate possible biases that could be introduced by the multiple test facilitators.

5.1 Ethics

All persons involved with the usability test are required to adhere to the following ethical guidelines:

- The performance of any test participant must not be individually attributable. Individual participant's name should not be used in reference outside the testing session. Toward this end, all participants are issued a unique identifyier number, and only the number is used to track collected data. No link between a participants name and his/her identifier number will be stored.

6. Usability Tasks

The usability tasks were derived from test scenarios developed with the assistance of members of the Geopod development project team, which includes experts in both earth science and computer science. Due to the range and extent of functionality provided in the application, and the short time for which each participant will be available, the tasks are the most common and relatively complex of available functions. The tasks are identical for all participants in the study.

The four test trials (Appendix VII) were of two types. The first 3 trials were composed by combining "base" tasks (tasks that accomplish a short specific goal) suggested by the application developers. These base tasks (Appendix VIII) were combined into 3 short scenarios of between 15 and 20 steps, with anticipated task times of approximately 3 to 5 minutes. These trials are meant to test the various controls available to the user, and to demonstrate the intuitive nature of the interface. Little attempt is being made to make the tasks of the first three trials mimic real activities; although the tasks given are representative of those that might be used in a typical student assignment, no attempt was made to give these tasks an earth science student assignment context. Rather, the emphasis is on the participant using each interface control multiple times in order to observe any potential interface problems with those controls. The first 3 trials would use a simple data set that displays a single "isosurface" for the participants to explore.

The fourth trial, however, was composed by an earth science faculty member to resemble a typical student assignment, intended to take approximately 15 to 20 minutes to complete. Trial 4 involves 3 isosurfaces in a complex configuration. This trial is meant to be a more realistic exercise of the system, and will be used to determine any difficulties caused by a more complex environment and a significantly longer use. There is an assumption that participants will learn from the first three trials, and will be able to successfully put that learning to use during this final trial.

The Geopod software has been modified to include a "Done" button, which will be used by the participant to signal that they have finished all of the tasks for a given trial. This modification does not in any way impact the performance of the system, and the button is located in an out-of-the-way place on the screen in order to minimize the risk that the participant will accidently activate it. Since the effect of clicking the "Done" button is to simply minimize the Geopod window to the system task tray without loss of data, the session could be resumed (and the time to task completion adjusted to account for the interval) should the user accidently click this button.

7. Usability Metrics

Usability metrics refers to user performance measured against specific performance goals necessary to satisfy usability requirements. Scenario (task) completion success rates, adherence to dialog scripts, error rates, and subjective evaluations will be used. Time-to-completion of tasks will also be collected.

7.1 Scenario Completion

Each trial requires the participant to obtain or input specific data that would be used in the course of a typical task. The trial is completed when the participant indicates that they have finished the task by clicking on the "Done" button of the software.

Participants are told (during the facilitator's testing protocol) that, should they not be able to figure out how to complete a step in a trial, they should verbally announce "Skipping step" and then the number of the step skipped.

Independent completion of the scenario is a universal goal. At no time should the test facilitators intervene with a given task.

7.2 Critical Errors

Critical errors are deviations at completion from the targets of the scenario. Obtaining or otherwise reporting of the wrong data value due to participant workflow is a critical error. Participants may or may not be aware that the task goal is incorrect or incomplete.

Critical errors can also be assigned when the participant initiates (or attempts to initiate) an action that will result in the goal state becoming unobtainable. In general, critical errors are unresolved errors during the process of completing the task or errors that produce an incorrect outcome.

7.3 Non-critical Errors

Non-critical errors are errors that are recovered from by the participant or, if not detected, do not result in processing problems or unexpected results. Although non-critical errors can be undetected by the participant, when they are detected they are generally frustrating to the participant.

These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (e.g., excessive steps and keystrokes). These errors may also be errors of confusion (ex., initially selecting the wrong function, using a user-interface control incorrectly selecting the wrong control or tool).

Noncritical errors can always be recovered from during the process of completing the scenario. Exploratory behavior, such as opening the wrong menu while searching for a function, will be coded as a non-critical error.

7.4 Subjective Evaluations

Subjective evaluations regarding ease of use and satisfaction will be collected via questionnaires, and during debriefing at the conclusion of the session. The questionnaires will utilize free-form responses and rating scales.

7.5 Scenario Completion Time (time on task)

The time to complete each scenario, not including subjective evaluation durations, will be recorded.

7. Usability Goals

The next section describes the usability goals for the Geopod project.

7.1 Completion Rate

Completion rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct.

A completion rate of 100% is the goal for each task in this usability test.

7.2 Error-free rate

Error-free rate is the percentage of test participants who complete the task without any errors (critical **or** non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

An error-free rate of [80%/enter error-free rate] is the goal for each task in this usability test.

7.3 Time on Task (TOT)

The time to complete a scenario is referred to as "time on task". It is measured from the time the person begins the scenario to the time he/she signals completion.

The goal time on task for each trial is

- Trials 1-3: 4 minutes
- Trial 4: 20 minutes

7.4 Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/debriefing session, these data are used to assess attitudes of the participants. See Appendix VI for the survey instrument used.

8. Problem Severity

To prioritize recommendations, a method of problem severity classification will be used in the analysis of the data collected during evaluation activities. The approach treats problem severity as a combination of two factors - the impact of the problem and the frequency of users experiencing the problem during the evaluation.

8.1 Impact

Impact is the ranking of the consequences of the problem by defining the level of impact that the problem has on successful task completion. There are three levels of impact:

- High - prevents the user from completing the task (critical error)

- Moderate - causes user difficulty but the task can be completed (non-critical error)
- Low - minor problems that do not significantly affect the task completion (non-critical error)

8.2 Frequency

Frequency is the percentage of participants who experience the problem when working on a task.

- High: 30% or more of the participants experience the problem
- Moderate: 11% - 29% of participants experience the problem
- Low: 10% or fewer of the participants experience the problem

Problem Severity Classification

The identified severity for each problem implies a general reward for resolving it, and a general risk for not addressing it, in the current release.

Severity 1 - High impact problems that often prevent a user from correctly completing a task. They occur in varying frequency and are characteristic of calls to the Help Desk. Reward for resolution is typically exhibited in fewer Help Desk calls and reduced redevelopment costs.

Severity 2 - Moderate to high frequency problems with moderate to low impact are typical of erroneous actions that the participant recognizes needs to be undone. Reward for resolution is typically exhibited in reduced time on task and decreased training costs.

Severity 3 - Either moderate problems with low frequency or low problems with moderate frequency; these are minor annoyance problems faced by a number of participants. Reward for resolution is typically exhibited in reduced time on task and increased data integrity.

Severity 4 - Low impact problems faced by few participants; there is low risk to not resolving these problems. Reward for resolution is typically exhibited in increased user satisfaction.

9. Reporting Results

The Usability Test Report will be provided at the conclusion of the usability test. It will consist of a report and/or a presentation of the results; evaluate the usability metrics against the pre-approved goals, subjective evaluations, and specific usability problems and recommendations for resolution. The recommendations will be categorically sized by development to aid in implementation strategy. The report is anticipated to be delivered to the Project team by Jan 1, 2011.

[i] Nielson, J. Usability engineering at a discount. 3rd International Conference on Human-Computer Interaction. Boston, MA.18-22 September, 1989.