AutoDrive

Usability Test Plan

Version 1.1

Jianjie Liu, Michael Morscher
February 23, 2019

Table of Contents

Document Overview

This document describes a test plan for conducting a usability test during the development of AutoDrive. 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 usability of the software.

The usability test objectives are:

- To determine design inconsistencies and usability problem areas within the user interface and content areas. Potential sources of error may include:
  - Navigation errors – failure to locate desired function.
  - Presentation errors – failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities.
  - Control usage problems – software crashes due to improper command handling.
- Exercise the application controlled test conditions with representative users. Data will be used to access whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.
- Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

The targeted user group should be anyone with minimum software engineering background that understands how to use command terminal in a machine running Linux.

Executive Summary

Users should interact with our software through running a Python script on a Linux command terminal. Our tests should be focused on how closely resembling is our command line interface with the traditional Linux terminal commands in terms of the presentation and description of the script's keywords when user append "-h" or "-help" when running our Python script.

Upon review of this usability test plan, including the draft task scenarios and usability goals for AutoDrive, documented acceptance of the plan is expected.

Methodology

We will recruit 5 participants from various technological depth and give them access to a Linux system with our software installed. We will ask them to actuate our self-driving toy car solely through reading the provided command line manual. The participant should give feedback on the content of the manual and what desire functions may be missing.

Participants

We will recruit 4 student participants of various grade level from Freshman to Senior majoring in computer science and 1 professor from the computer science department.

The participants' responsibilities will be to attempt to complete a set of representative task scenarios presented to them in as efficient and timely a manner as possible, and to provide 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 and debriefing.

Training
No training is necessary in this usability test.

Procedure
Participants will take part in the usability test at room 122 in Halligan Hall. A laptop computer with our software installed will be used in a typical office environment. The participant's interaction with the software will be monitored by the facilitator seated in the same office. Note takers and data logger(s) will monitor the sessions in observation room.

The facilitator will brief the participants on the application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant. We will ask the participants a verbal consent that they acknowledge the participation is voluntary which they can cease at any time, and any data or description produced from the session will remain anonymous. The facilitator will ask the participant if they have any questions.

The facilitator will observe and enter user behavior, user comments, and system actions on an excel sheet.

After each task, the participant will complete the post-task questionnaire and elaborate on the task session with the facilitator. After all task scenarios are attempted, the participant will complete the post-test satisfaction questionnaire.

Roles
The roles involved in a usability test are as follows. An individual may play multiple roles and tests may not require all roles.

Trainer
- Provide training overview prior to usability testing

Facilitator
- Provides overview of study to participants
- Defines usability and purpose of usability testing to participants
- Assists in conduct of participant and observer debriefing sessions
- Responds to participant's requests for assistance

Data Logger
- Records participant's actions and comments

Test Observers

- Silent observer
- Assists the data logger in identifying problems, concerns, coding bugs, and procedural errors
- Serve as note takers.

Test Participants

- Provides overview of study to participants
- Defines usability and purpose of usability testing to participants
- Assists in conduct of participant and observer debriefing sessions
- Responds to participant's requests for assistance

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.
- A description of the participant's performance should not be reported to his or her manager.

Usability Tasks

Participant will be asked to read the manual of our Python script and enter the corresponding commands on the terminal to actuate the car for

1) fixed amount of time
2) and indefinitely.

If participant can successfully actuate the car through entering the correct command within 10 minutes since the start of the test, then the participant has successfully completed the scenario.

Usability Metrics

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

Scenario Completion

Each scenario will require, or request, that the participant obtains or inputs specific data that would be used in course of a typical task. The scenario is completed when the participant indicates the scenario's goal has been obtained (whether successfully or unsuccessfully) or the participant requests and receives sufficient guidance as to warrant scoring the scenario as a critical error.

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.

Independent completion of the scenario is a universal goal; help obtained from the other usability test roles is cause to score the scenario a critical error. Critical errors can also be assigned when the participant initiates (or attempts to initiate) and 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.

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 such as attempting to edit an un-editable field).

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 not be coded as a non-critical error.

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.

Scenario Completion Time (time on task)

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

Usability Goals

The next section describes the usability goals for AutoDrive.

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. Note: If a participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

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

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 70% is the goal for each task in this usability test.

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.

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.

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.

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)

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.

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.