

Usability Evaluation of High-Level User Assistance for Robot Mission Specification

Yoichiro Endo, Douglas C. MacKenzie, and Ronald C. Arkin

Abstract—*MissionLab* is a mission specification system that implements a hybrid deliberative and reactive control architecture for autonomous mobile robots. The user creates and executes the robot mission plans through its graphical user interface. As robot deployments become more common in highly stressful situations, such as in dealing with explosives or biohazards, the usability of their mission specification system becomes critical. To address this need, a mission-planning “wizard” has been recently integrated into *MissionLab*. By retrieving and adapting past successful mission plans stored in its database, this new feature is designed to simplify the user’s planning process. The latest formal usability experiments, reported in this paper, testing for usability improvements in terms of speed of the mission planning process, accuracy of the produced mission plans, and ease of use is conducted. This paper introduces the mission-planning wizard, describes the usability experiments (including design), and discusses the results in detail.

Index Terms—Usability Evaluation, Robot Mission Specification System.

I. INTRODUCTION

For the past nine years, Georgia Tech has been developing a robot mission specification system, called *MissionLab*, which allows users to design and execute autonomous mobile robot missions using a graphical user interface (GUI) [5,19]. *MissionLab* supports planning of robot missions suitable for military applications, including explosive ordinance disposal (EOD), indoor assessment missions (such as locating biohazards), and outdoor waypoint navigation [8,14]. According to recent surveys [13,21], these types of robotic missions are also useful for civilian law enforcement applications. As more robot mission plans are composed by end-users who are not computer programmers and who work in highly stressful situations, we expect that the usability of the mission specification system will become crucial. In order to ensure the acceptance of *MissionLab* by users and assess improvements in its usability, we have repeatedly conducted formal usability studies examining its user interface. This paper presents the results of our most recent study, conducted within DARPA’s Mobile Autonomous Robot Software (MARS) program.

Formal usability studies require preparation of an experiment, execution of the experiment following strict protocols, and analysis of the gathered data. For example, in Adams’ Ph.D. dissertation [1] (one of the earliest formal usability studies conducted on robotic software we know of), the user interface of her multiagent robotics system was evaluated with a formal usability experiment, and a comprehensive analysis of the experimental data was presented to show how such an evaluation could be conducted formally in the robotics domain.

We believe that the *MissionLab* is one of the most extensively studied robotics toolset in terms of the usability. For example, MacKenzie and Arkin [18] evaluated the *MissionLab* GUI by contrasting it with a traditional non-GUI method (text editing) of mission planning. Similarly, the work by Ali [2] evaluated the usability of teleautonomous control interfaces within *MissionLab* from a multiagent robotics perspective. Subsequently, as part of DARPA’s Tactical Mobile Robotics (TMR) program, another study [14] conducted on *MissionLab* evaluated the usability of its mission planning process to compose tactical mission plans suitable for urban warfare. The goal of the latest usability study was to assess a new mission-planning “wizard” feature in *MissionLab*. This new interface is designed to assist the user by suggesting a mission plan based on previous successful plans stored in a database.

Several other examples of usability studies in robotics can be found in the rehabilitation robotics domain. Simpson et al. [24], for example, had conducted a usability experiment using eight subjects in order to determine the utility of their assistive wheelchair navigation system. While the number of the subjects who represented the true target end-users (i.e., ones with physical disabilities) was rather small (only four), nevertheless, they were successfully able to identify the limitations in the current design and ways to improve their system. Likewise, Buhler et al.[6], Parsons et al. [22], and Keates et al. [17] each sought to assess the effectiveness of their robotic manipulator interfaces using target end-users (patients with physical disabilities). However, their experiments seem to have been loosely defined, leaving the validities of their results uncertain. Similarly, while Montemerlo et al. [20] reported that their autonomous nursing robot was able to guide the target end-users (elderly people) in the indoor environment effectively, in our opinion the lack of statistical validation due to improper setup of the usability testing left their claim rather weak.

On the other hand, solid robotics human factor studies can be found for teleoperation in nuclear environments [9,10,11] although these focus on run-time control and are not closely related to the research described in this article.

This research is supported by DARPA U.S. Army SMDC contract #DASG60-99-C-0081. Approved for Public Release, distribution unlimited.

Yoichiro Endo and Ronald C. Arkin are with Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, Atlanta, Georgia 30332-0280, USA (email: {endo, arkin}@cc.gatech.edu). Douglas C. MacKenzie is with Mobile Intelligence Corporation, 33150 Schoolcraft Road, Suite 108, Livonia, Michigan 48150-1646, USA (email: doug@mobile-intelligence.com).

An overview of *MissionLab* and the new mission-planning wizard are presented in Section II. A description of our latest usability study, including preparation, execution, and analysis of the data gathered is described in Section III. Finally, Section IV presents our conclusions.

II. MISSIONLAB

This section presents an overview of the existing *MissionLab* mission specification capabilities (called the “Basic Features”) and the new mission-planning wizard. *MissionLab* embodies a hybrid deliberative and reactive control system as defined by the Autonomous Robot Architecture (AuRA) [3,4]. It runs under RedHat Linux, and can control various commercial robots, including ATRV-Jr (iRobot), AmigoBot (ActivMedia), Nomad 200 (Nomadic Technologies), and many others. The software is freely available on the Georgia Tech Mobile Robot Laboratory website¹.

A. Mission Planning Using MissionLab’s Basic Features

In *MissionLab*, the user composes a mission plan using a set of finite state acceptors (FSAs) (details on the use of FSAs in this context can be found in [3]). The user-specified plan consists of tasks (behavioral states) and perceptual triggers (transition functions). A task is an assemblage of behaviors that the robot executes, and a perceptual trigger is a condition that, when true, causes a state transition to the subsequent task. The user creates a mission by choosing an appropriate sequencing of tasks by adding perceptual triggers to control the order of execution.

FSA-based mission composition is performed using the *CfgEdit* (Configuration Editor) module. Figure 1 depicts an example FSA mission representation created in *CfgEdit*. The tasks appear as circles and the perceptual triggers are represented as arrows. The mission shown is an example of an explosive ordnance disposal (EOD) mission, where *Start*, *LookFor*, *MoveToward*, *PickUp*, *PutInEOD*, and *Terminate* are tasks, and *Immediate*, *Detect*, *Near*, *Holding*, *NotHolding*, and *NotDetected* are the perceptual triggers. In this mission, the robot immediately looks for a mine, and once it is detected, the robot moves towards the mine, picks it up, then carries it to an EOD area, and disposes of it there. If no mines are detectable, the robot returns to its home base and terminates the mission.

The “waypoints specification” tool is another basic feature that has been implemented in *MissionLab*. The user places a waypoint on a map-like representation of the area called an “overlay”. The coordinates of these waypoints are then translated into a series of (*GoTo*, *Near*) FSA state-transition pairs. These *GoTo* tasks command the robot to move towards the specified waypoint coordinates and when the robot reaches the waypoint the *Near* trigger activates, causing a transition to the next task.

Even though the process of creating a mission plan is assisted by the *MissionLab* GUI, it still requires the user to laboriously specify all aspects of the mission. The new

mission-planning wizard was designed to reduce this burden using case-based design techniques.

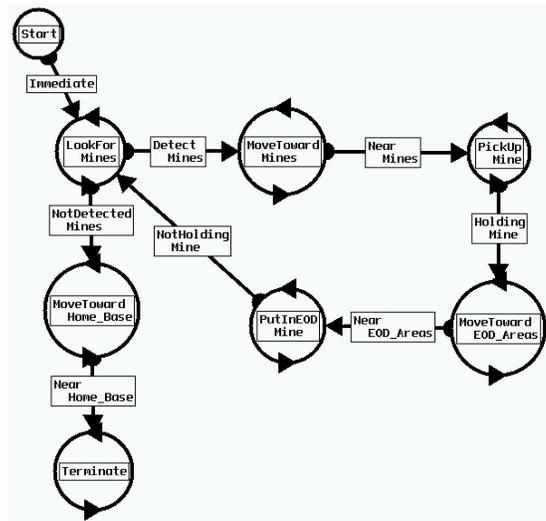


Figure 1: Example FSA for explosive ordnance disposal (EOD) mission.

B. The Mission-Planning Wizard

Recently, a new mission-planning “wizard” feature was created and integrated into the *MissionLab* system. This tool was designed to help users build mission plans by retrieving and adapting previous successful missions to fit the new situation. Its design was especially intended to benefit users who repeatedly create complex, but somewhat similar missions. The evaluation of the mission-planning wizard’s usability is the focus of the experimental study described in the remainder of this article.

The mission-planning wizard supports the user’s mission specification process by suggesting a plan that potentially meets the user’s intention. As shown in Figure 2, past successful missions are stored in the database of *miCBR*, a Java-based program developed by Mobile Intelligence Corporation. In *miCBR*, a previous successful mission, described in Configuration Description Language (CDL) [15,19] is stored as a “case” along with indexing features (requirements and preferences), unique identifiers, and its quantitative status (Figure 3). The user specifies the following requirements and preferences in order to specify a plan using the mission-planning wizard:

- **Task Category:** Defines the high-level task that should be performed in this segment of the mission. Note that this high-level “task” is different from the behavioral-state task mentioned in conjunction with FSA-based mission plans in Section II.A. This high-level task is a segment of the FSA-based mission, and a particular implementation can contain a complete FSA with its own behavioral-state tasks and perceptual triggers. The task categories used in the usability experiments include *Biohazard*, *EOD*, and *Waypoints*.
- **Environment:** Specifies whether the high-level task will be conducted inside a building or outdoors.

¹ <http://www.cc.gatech.edu/ai/robot-lab/research/MissionLab>

- **Localization Method:** States the user’s preference for using maps (GPS, gyroscope, etc.) or perceivable landmarks (color cues, object shapes, etc) for navigation.
- **Enemy Consideration:** Indicates whether the possible existence of an enemy agent should be considered during planning.
- **Number of Robots:** Indicates how many robots are involved in the overall mission.
- **Maximum Velocity:** Specifies the upper speed limit that the robot is allowed to travel.
- **Aggressiveness:** Specifies the suggested average speed of the robot.
- **Starting Coordinates:** Specifies the robot’s starting position for this segment.

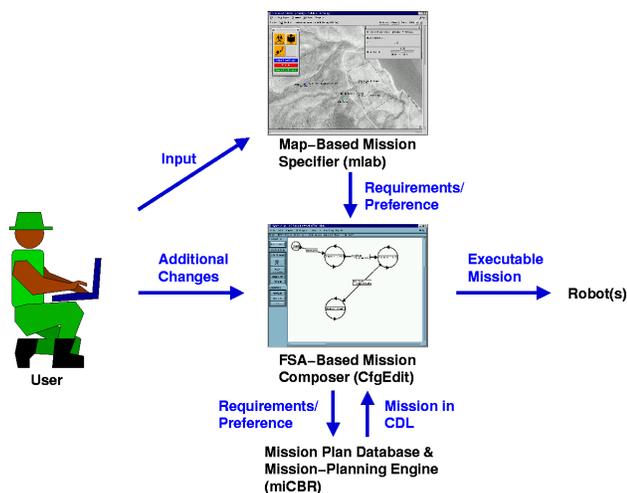


Figure 2: Mission specification using the *MissionLab* system with the mission-planning wizard feature. The user plans a mission on an overlay by specifying the requirements and preferences of the mission. Based on the requirements and preferences, the mission-planning engine in *miCBR* then suggests a plan from stored past successful mission plans within the database. The user can modify the suggested FSA-based plan if so desired. The resulting mission plan can then be compiled, and executed on a simulated or real robot.

TaskCategory:	EOD	CaseID:	237
Environment:	Outdoor	Quality:	0.75
LocalizationMethod:	UseLandmarks		...
EnemyConsideration:	False		
NumberOfRobots:	1		
...			
CaseData: bindArch AuRa.urban; InsBP \$AN_1403 from movement(base_vel = {0.1}, bound_to = base:DRIVE_W_SPIN(v = FSA (...			

Figure 3: Example of a mission plan stored in the *miCBR* database. The case features (requirements and preferences) appear in the upper-left, identifiers and status in the upper-right, and the case data (a mission plan described in CDL) are shown in the bottom. This case is relevant when a user tries to create a mission where one robot conducts an EOD task outdoors and the exact location of the mines are not available, and an enemy agent is not likely to be present.

The user specifies these mission requirements and preferences using the *mlab* map-interface module. Figure 4 shows an example where the user is creating a mission in which a robot conducts an EOD task at the “Expected Refugee Camp Site” location, after first following two waypoints from the initial start position (labeled “StartPlace”). The user first has to specify the two “waypoint” tasks by setting the *Task Category*. *Task Category* is selected using the *Waypoints* icon button from the toolbox window (upper left). Once the user places the two waypoints on the field, he/she next places an *EOD* task. The clicked map locations are automatically translated to *Starting Coordinates*. Selecting the *Number of Robots* and the *Maximum Velocity* is done in the “Mission Preference” window (upper right), and the rest of the requirements and preferences (*Environment*, *Localization Method*, *Enemy Consideration*, and *Aggressiveness*) are set in the task-dependent “Task Preference” window (lower left).

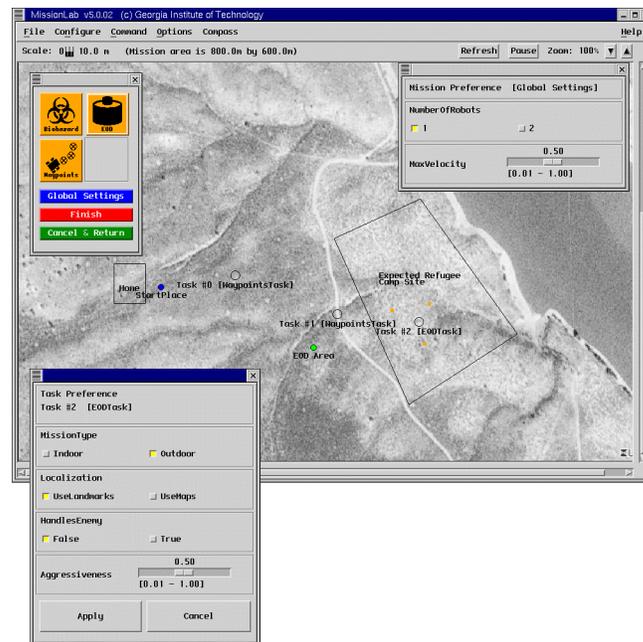


Figure 4: *mlab* as an interface of the mission-planning wizard.

After the mission requirements and preferences are specified, the information is passed to the *miCBR* module. The mission-planning engine of *miCBR* then constructs a mission plan by retrieving and adapting the mission plans that are closest to the requirements and preferences of the desired mission. For instance, in the case of the waypoints-EOD mission example mentioned above, if the requirements and preferences shown in Table 1 are sent to the *miCBR* module, it will return a mission plan similar to the one shown in Figure 5.

III. USABILITY STUDY

The reason for the usability study was to determine whether the mission-planning wizard actually improved the usability of *MissionLab* in terms of the user’s mission planning experience. In this section, both preparation of the usability experiment, and the results of the experiment are explained.

A. Preparing the Usability Experiment

The flow chart depicted in Figure 6 illustrates our recommended procedure to follow when preparing usability experiments of this kind. In the figure, an arrow denotes a dependency of one block (step) on another; the step in the upper block has to be taken before the step in the lower block can be executed.

First, hypotheses of expected usability improvements in the system should be established. These hypotheses will allow the experimental designer to focus on relevant aspects of the usability, and then determine appropriate experimental methods to measure them. After the experimental methods are identified, the questions regarding subject demographics (e.g., “what kind of backgrounds the experiment participants should have?”) and the hardware setup (e.g., “what kind of equipment do we need?”) can be answered. Documentation necessary for the experiment, such as protocols, can then be prepared based on the determined experimental methods and the expected subject backgrounds. The experimental software can be prepared once the system hardware is configured. Finally, we recommend recruitment of the subjects to be started only when the expected duration of the experiment is known. Such estimation, however, may be only available after running a dry-run session using the complete protocols and software. Examples from our experiment are provided below.

Mission Plan	Requirements / Preferences	Value
1 st Segment	Task Category	Waypoints
	Environment	Outdoor
	Localization Method	UseMaps
	Enemy Consideration	False
	Number of Robots	1
	Maximum Velocity	0.5
	Aggressiveness	0.5
Starting Coordinates	301.01, 228.57	
2 nd Segment	Task Category	Waypoints
	Environment	Outdoor
	Localization Method	UseMaps
	Enemy Consideration	False
	Number of Robots	1
	Maximum Velocity	0.5
	Aggressiveness	0.5
Starting Coordinates	401.61, 197.18	
3 rd Segment	Task Category	EOD
	Environment	Outdoor
	Localization Method	UseLandmarks
	Enemy Consideration	False
	Number of Robots	1
	Maximum Velocity	0.5
	Aggressiveness	0.5
Starting Coordinates	499.80, 174.65	

Table 1: Example requirements and preferences for a Waypoints-EOD mission.

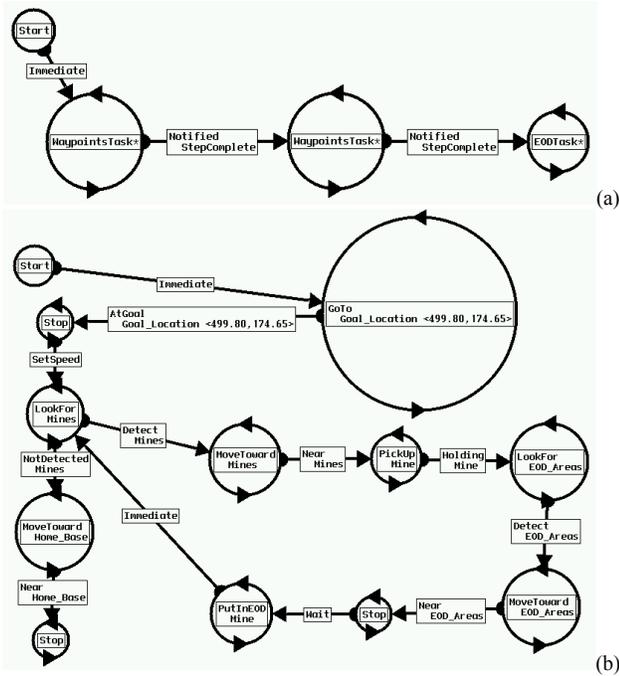


Figure 5: Mission plans suggested by the mission-planning wizard based on the example requirements and preferences shown in Table 1; (a) high level representation of the mission plan, showing two waypoints task, and an EOD task; (b) the nested sub-FSA for the EOD task in (a).

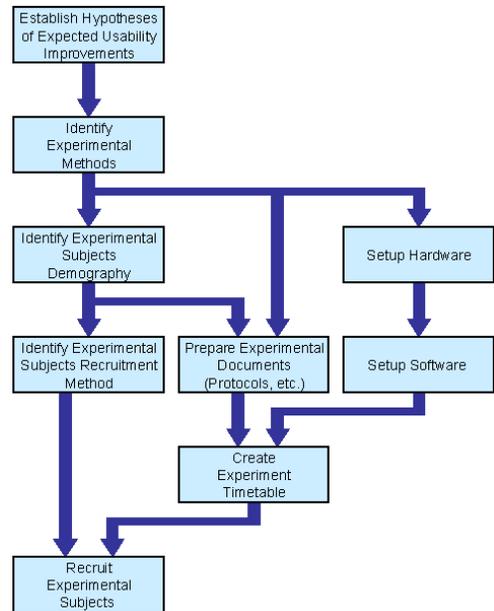


Figure 6: Recommended steps for preparing the usability experiment

1) Hypotheses

The following form the hypotheses of expected usability improvements, and are used to guide the experiment to measure quantitatively just how much, if at all, the mission-planning wizard improves the usability of *MissionLab*.

Hypothesis 1: The mission-planning wizard reduces the time required to create a suitable mission plan.

Hypothesis 2: The mission-planning wizard increases the accuracy of mission plans created by users.

Hypothesis 3: Users find it easier to create a mission plan with the mission-planning wizard than without it.

In other words, the usability experiment was specifically tailored to verify whether: (1) the amount of the time it takes during the mission planning process using the mission-planning wizard would be indeed less than using the system with just the basic features of *MissionLab*; (2) the robot running the mission plan created with the mission-planning wizard would have better performance than the one created without it (in terms of accomplishing the mission requirements for a particular class of missions); (3) users creating a mission plan using the mission-planning wizard would find *MissionLab* easier to use (in terms of overall usage, learning the functionality of the system, and finding appropriate buttons) than users who create mission plans using only the basic *MissionLab* features.

2) Experimental Methods

The above hypotheses were evaluated by comparing *MissionLab* systems with and without the mission-planning wizard. All of the experimental subjects went through a series of tutorials and two sets of tests. In Test 1, each subject was asked to create a realistic military mission (indoor assessment) for a single robot. The scenario is based on our real robot experiments reported in [8] and [14]. To add to the complexity to the mission, in Test 2, the subject was asked to create a similar indoor-assessment mission using two robots. This mission involved the coordination of two robots in the presence of an enemy agent.

To measure the effectiveness of the mission-planning wizard, the subjects were divided into two control groups. For the first group (Group A), the tests were carried out using the basic *MissionLab* features (i.e., no mission-planning wizard). The second group (Group B) was introduced to the mission-planning wizard during one of the tutorials, and, unlike Group A, the mission-planning wizard was available to the subject during the tests.

To evaluate Hypothesis 1, time-stamps of various keyboard and mouse events generated during the tests were recorded using the *MissionLab* logging features. The most important measurement was the total time required to create each mission plan. The ability of the mission-planning wizard to expedite the mission planning process was determined by comparing the time spent in the mission planning process between Group A and Group B.

Hypothesis 2 was evaluated by comparing the “accuracy” (success rates) of the mission plans created by the subjects; each mission plan was compiled and executed

by a *MissionLab* expert after the tests, and the accuracy of the mission was computed based on the mission-specific criteria (Table 2 and Table 3). For example, a robot executing the mission plan created for Test 1 would have to complete 10 specified tasks, while two robots for Test 2 would have finish 15 specified tasks. The percentage of the cleared tasks out of the total specified tasks was used as a metric of the accuracy.

In order to evaluate Hypothesis 3, all the subjects were asked to fill out a set of questionnaires before and after the tests. The pre-test questionnaire was designed to acquire background information about the subject, and the post-test questionnaire solicited their opinion on how easy it was to create mission plans using *MissionLab*.

Finally, 5 experiment administrators were recruited from Georgia Tech Mobile Robot Laboratory. The job of the administrators was to manage the usability experiments including subject training, and testing using predefined protocols. All of the administrators were *MissionLab* experts who had a year or more of experience developing the software.

Criterion #	Task to Be Cleared
1	Robot followed waypoints
2	Robot entered the building
3	Robot proceeded along the hallway
4	Robot entered a room
5	Robot surveyed a room
6	Robot alerted for biohazard
7	Robot marked the door
8	Robot entered more than one room
9	Robot turned back at end of the hall
10	Mission terminated after alerting

Table 2: Mission-specific criteria for Test 1 (Appendix A). The subject-generated mission plans were compiled and executed in simulation to examine the percentage of these criteria being met by the simulated missions. This percentage was considered to be the “success rate” of the mission.

Criterion #	Task to Be Cleared
1	Robot1 followed waypoints
2	Robot2 followed waypoints
3	Robot1 entered the building
4	Robot2 entered the building
5	Robot1 proceeded along hallway
6	Robot1 entered a room
7	Robot2 overwatched Robot1
8	Robot1 ducked from an enemy
9	Robot2 ducked from an enemy
10	Robot1 surveyed a room
11	Robot1 alerted for biohazard
12	Robot1 marked the door
13	Robot1 entered more than one room
14	Robot1 turned back at end of the hall
15	Mission terminated after alerting

Table 3: Mission-specific criteria for Test 2 (Appendix B).

3) Experimental Subjects

The experimental subjects who participated in this study were restricted to people who had never been introduced to *MissionLab* before. As mentioned above, the subjects were divided into two control groups: one group (Group A) who would only use the basic *MissionLab* features, and the other group (Group B) who would also be introduced to the

mission-planning wizard in one of the tutorials. A total of 29 subjects (14 subjects in Group A, and 15 subjects in Group B) participated in the experiments. Diverse subjects were recruited so that comparisons among subjects with different backgrounds could be analyzed (Figure 7). For example, 9 out of 14 subjects in Group A, and 11 out of 15 subjects in Group B had technical backgrounds. Here, any subject who either was working on a job related to, or was studying an area related to engineering, computer science, or math-related topics is considered to have a “technical background”.

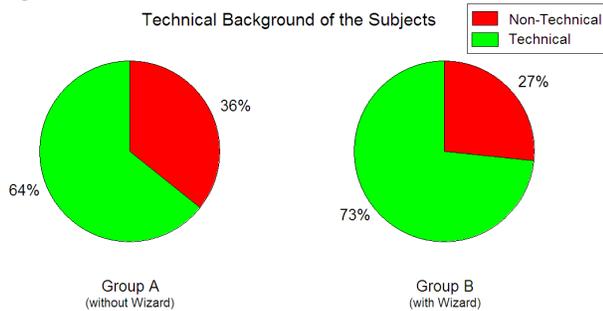


Figure 7: Technical demographics of the participating subjects (14 subjects are in Group A, and 15 subjects are in Group B). The two pie charts show the percentages of the subjects with technical backgrounds (engineering, computer science, or math) for each group.

Choosing how many subjects to use in a usability testing is a crucial but difficult task. While increasing the number of the subjects is likely to increase statistical significance, it also costs more and takes more time to complete the testing. Furthermore, different researchers recommend different numbers of subjects to be used for each experimental condition: e.g., “four or five” [26], “six or eight” [16], “at least 10 to 12” [23], and “10 to 20” [25]. In our case, using 14 to 15 subjects per group (per condition of having or not having the mission-planning wizard) reflects the higher end of the above spectrum. However, we were able to recruit only 4 to 5 subjects for each group who had no technical background, which we consider a relatively small sampling size.

4) Experimental Documents

Brief descriptions of all the documents used during the experiments follow; the first two documents are forms filled out by the subjects, and the remaining documents are experiment protocols. While copies of the protocols for Test 1 and Test 2 can be found in the Appendix, readers who wish to review these documents in detail should refer to [12].

- **Consent to Be a Research Participant:** This is a legal form required by the Georgia Tech Institutional Review Board. The purpose of the experiment and the subject’s rights during the experiment are explained in this document. The experiment cannot be conducted unless the subject signs this form.
- **Questionnaire:** The questionnaire serves two purposes: (1) to acquire the subject’s background information before the test, and (2) to obtain opinions after the test

on how easy or difficult it was to create mission plans using *MissionLab*.

- **Tutorial 1: Back and Forth Mission:** This protocol is for the first tutorial and is designed to provide the subject with a basic introduction to *MissionLab*. In this tutorial, the subject is asked to create a simple robot mission, in which the robot simply travels between two points.
- **Tutorial 2: Approach CoC Mission:** This tutorial was designed to teach the subject how to specify waypoints using the map interface. The subject is asked to construct a mission plan where the robot travels from one location to another using waypoints placed upon an overlay by the subject.
- **Tutorial 3: EOD Mission:** This tutorial teaches the subject how to create complex FSA mission plans. The subject is asked to create a mission plan causing a robot to collect explosives from a minefield and dispose of them in a safe area.
- **Tutorial 4: Robot Communication:** This tutorial was designed to teach the subject how to create a mission plan for multiple robots, and coordinate the activities of multiple robots. Here, the subject learns these techniques by creating a simple two-robot mission plan in which one robot calls another robot to approach it, and then dismisses it after a few seconds.
- **Tutorial 5: Advanced EOD Mission:** This tutorial was provided only for subjects in Group B. It was designed to teach the subject how to use the mission-planning wizard to build mission plans. The mission-planning wizard is explained to the subject, and then he/she is asked to use it to create two EOD missions for first a single-robot and then a double-robot scenario using a specified overlay.
- **Test 1: Biohazard Mission:** This protocol is for Test 1, and the identical protocol was used for both subject groups. In this test, the subject is asked to create a robot mission plan in which the robot approaches a building by following a series of waypoints; then once it is inside the building, it proceeds along the hallway and searches every open room to check if biohazards are present. Each subject is limited to no more than 45 minutes to create the mission plan. Hard copies of the instructions (Appendix A), a list of available tasks, a list of available triggers, and the screenshots of the specified overlay are also provided to the subjects during the test.
- **Test 2: Biohazard Mission Part II:** This is a protocol for Test 2, where the subject is asked to create a mission plan similar to Test 1. The mission has to be planned, however, so that one robot watches over the other robot during the search for biohazards. As in Test 1, the duration of this test is 45 minutes. Hard copies of the instructions (Appendix B), a list of available tasks and triggers, a screenshot of a sample FSA for a biohazard mission for Test 1, and the specified overlay were provided to the subject during this test.

5) Hardware and Software Setups

The experimental equipment was set up in a quiet small office (150-ft²). A Dell Latitude laptop computer (a single 267 MHz Intel Pentium II processor with 94 MB RAM) was placed in a docking station with a separate monitor (ViewSonic 22”), keyboard, and three-button mouse.

A version of *MissionLab*, capable of both enabling and disabling the mission-wizard feature, was installed on the laptop computer. In order to ease the process of analyzing the data, all the subject logs were stored systematically in separate directories; indexed by the date, subject number, and type of tests.

Item	Duration
Consent to Be a Research Participant, and Subject Contact Information	3 min.
Questionnaire (Pre-Test)	3 min.
Tutorial 1: Back and Forth Mission	15 min.
Tutorial 2: Approach CoC Mission	15 min.
Tutorial 3: EOD Mission	20 min.
Break for the subject.	10 min.
Tutorial 4: Robot Communication	20 min.
Test 1: Biohazard Mission	3 min. Explanation + 45 min. Actual Test
Break for the subject. Administrator checks the video/audio equipment.	12 min.
Test 2: Biohazard Mission Part II	3 min. Explanation + 45 min. Actual Test
Questionnaire (Post-Test)	12 min.
Total	3 hours & 30 min.

Table 4: Timetable for the Group A (without the mission-planning wizard) experiment.

Item	Duration
Consent to Be a Research Participant, and Subject Contact Information	3 min.
Questionnaire (Pre-Test)	3 min.
Tutorial 1: Back and Forth Mission	15 min.
Tutorial 2: Approach CoC Mission	15 min.
Tutorial 3: EOD Mission	20 min.
Break for the subject.	10 min.
Tutorial 4: Robot Communication	20 min.
Tutorial 5: Advanced EOD Mission	20 min.
Break for the subject.	10 min.
Test 1: Biohazard Mission	3 min. Explanation + 45 min. Actual Test
Break for the subject. Administrator checks the video/audio equipment.	12 min.
Test 2: Biohazard Mission Part II	3 min. Explanation + 45 min. Actual Test
Questionnaire (Post-Test)	12 min.
Total	4 hours

Table 5: Timetable for the Group B (with the mission-planning wizard) experiment.

6) Experiment Timetable

The maximum duration of each session for Group A was 3 hours and 30 minutes, while it was 4 hours for Group B because of their extra tutorial. The tutorials, tests, and other items, as well as the designated duration for those items are shown in Table 4 and Table 5. Each of the two tests was assigned a period of 45 minutes. However, if a subject finished planning the mission and was satisfied with the simulation results, the test was concluded. The 45-

minute testing period was chosen based on the average time (45 minutes and 2 seconds) required by subjects creating a similar two-robot mission in the previous TMR usability study [14]. While, in the TMR usability study, two tests were conducted on two separate days for each subject, both tests were conducted during a single-day session in this study. This allowed us to reduce the total days of the experiment drastically (from 133 days to 79 days) while the total number of the subjects remained the same (29 subjects).

B. Results

Analysis of the data gathered during the usability experiment will now be presented. The three hypotheses introduced earlier in Section III.A were analyzed using a one-way analysis of variance (ANOVA) test and the results are discussed. The section concludes with a look at the correlation between the mission plan accuracy and the time expended by the subjects.

1) Hypothesis 1

The average time subjects took to create the mission plans was compared to evaluate Hypothesis 1: “the mission-planning wizard reduces the time required to create a suitable mission plan.” Measurement of each test duration was started when the subject began working with *MissionLab*, and ended when one of the two conditions were met: (a) the subject was satisfied with the simulation results of the composed mission plan; or (b) the 45 minutes assigned for the test had elapsed. The compilation and simulation times were subtracted from the total duration because the subject could not actively create the mission plan during those periods.

The graph in Figure 8 shows the means of the two test durations for each group. The vertical bars in the graph denote the 95% confidence intervals, which were computed by statistical software: *STATISTICA* (version 6.0; StatSoft, Inc.). *STATISTICA* was also utilized to conduct one-way ANOVA in order to determine the statistical significance of the differences among the means. We follow standard practice and consider a difference to be significant when the *p*-value of the associated ANOVA test is less than 0.05 (5%). All the statistical hypotheses, corresponding ANOVA tables, and the results of the significance testing conducted to analyze the data mentioned in this section are presented in Appendix C.

In Test 2, the mission-planning wizard reduced the time required to create missions, with the mean of Group B’s Test 2 duration (with the mission-planning wizard) significantly less than Group A’s ($p = 0.020$). In Test 1, however, while the mean of Group B’s test duration is less than Group A’s, the difference was not statistically significant ($p = 0.095$).

2) Hypothesis 2

The mean accuracies of the mission plans created by the subjects were compared to evaluate the hypothesis that “the

mission-planning wizard increases the accuracy of mission plans created by users” (Hypothesis 2). As shown in Figure 9 for Test 1, no notable difference between the mean accuracies for the two groups was found (the p -value for the variability of two group-means on Test 1 was 0.530). However, for Test 2, the mission-planning wizard did improve the accuracy of the mission plans, as the mean success rate of Group B (with the mission-planning wizard) was significantly higher than Group A’s ($p = 0.023$).

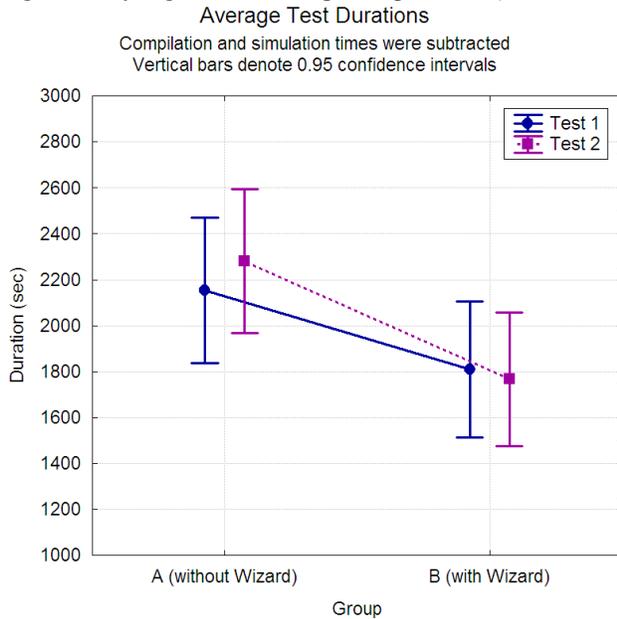


Figure 8: Comparison of average test duration between the two groups for both tests. In Test 2, Group A took a relatively longer time to plan the missions than did Group B ($p = 0.020$). However, for Test 1, a statistically significant difference between the two groups was not observed ($p = 0.095$).

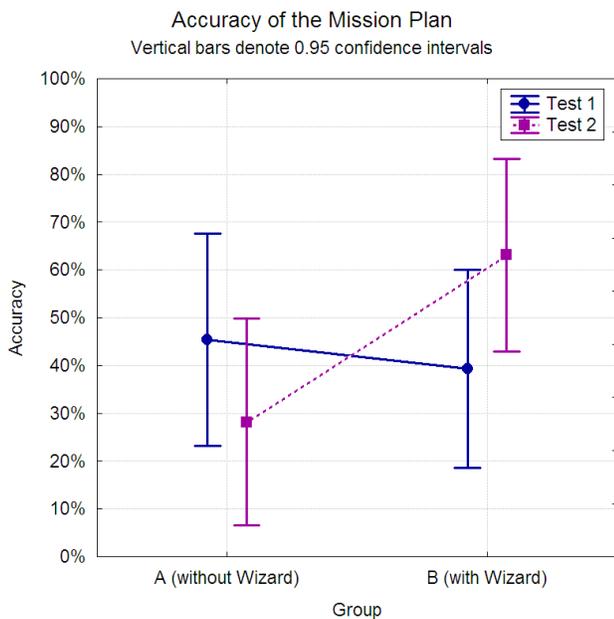


Figure 9: Comparison of the mean accuracies of the missions created for the two tests. Group B scored significantly higher success rate ($p = 0.023$) comparing to Group A in Test 2. However, for Test 1, a statistically significant difference between the two groups was not observed ($p = 0.530$).

3) Hypothesis 3

Hypothesis 3, “users find it easier to create a mission plan with the mission-planning wizard than without it,” was evaluated by analyzing feedback provided by subjects after the tests. They were asked to grade the difficulty of *MissionLab* using the scale 1 (very easy) to 5 (very difficult) in each of the four categories: (1) overall usage, (2) learning functionalities, (3) creating robot missions, and (4) finding buttons. The scores of the four difficulties were averaged to determine the general difficulty of *MissionLab* expressed by each subject. The graph in Figure 10 shows the means of the general difficulties depending on the group. As shown in the graph, the availability of the mission-planning wizard did not have a statistically significant impact on the general difficulties (the p -value for the variability of the two group-means was 0.438).

However, analyzing the group separately, a one-way ANOVA with respect to the technical background of the subjects revealed an interesting trend in terms of the general difficulty of *MissionLab*. For example, in Figure 11, the mean of the general *MissionLab* difficulty expressed by non-technical background subjects in Group A (without the mission-planning wizard) was substantially higher than those with technical background in Group A ($p = 0.005$). On the other hand, as shown in Figure 12, if the mission-planning wizard was available to the subjects (i.e., Group B), there was no substantial difference between the means of the general *MissionLab* difficulties expressed by those with a technical background and those without it ($p = 0.441$). In this case, the mission-planning wizard has compensated for the subjects’ lack of technical background. However, it should be noted that the sampling size used in this analysis was relatively small, as only 5 out of 14 subjects in Group A and 4 out of 15 subjects in Group B had non-technical backgrounds.

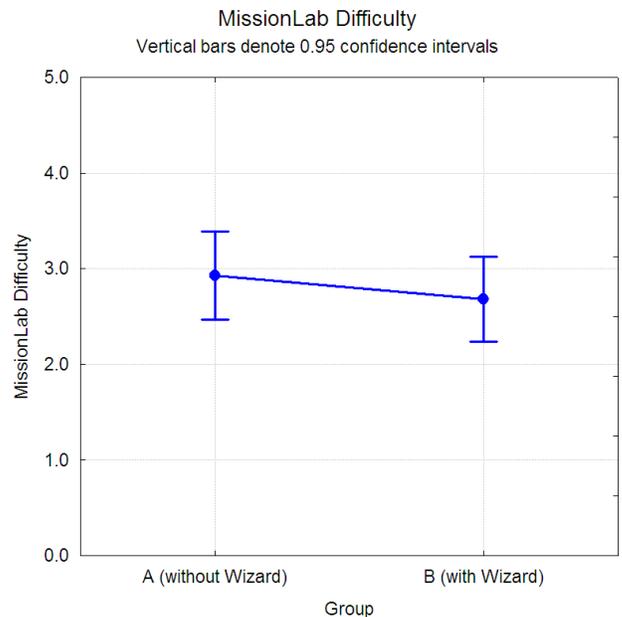


Figure 10: Comparison of the general *MissionLab* difficulty means between two groups. No significant difference between the two group-means was found ($p = 0.438$).

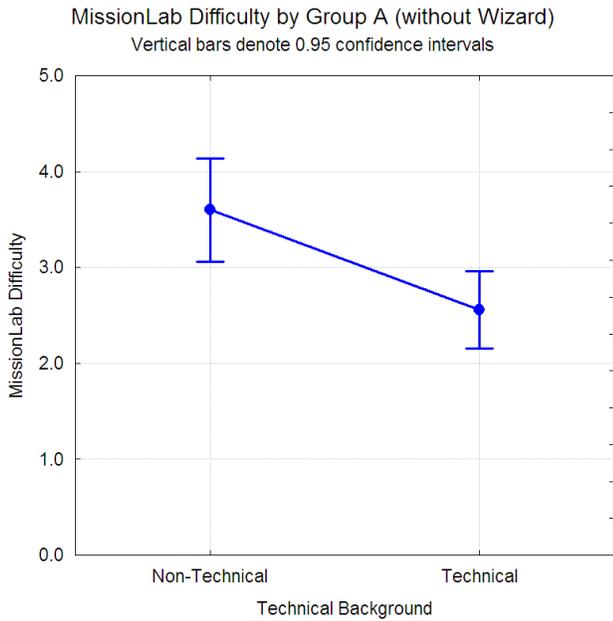


Figure 11: Comparison of the general *MissionLab* difficulty means between the Group A subjects (without the mission-planning wizard) with and without technical backgrounds. The mean of the general difficulty expressed by the Group A subjects without the technical background was substantially higher than those with the technical background ($p = 0.005$).

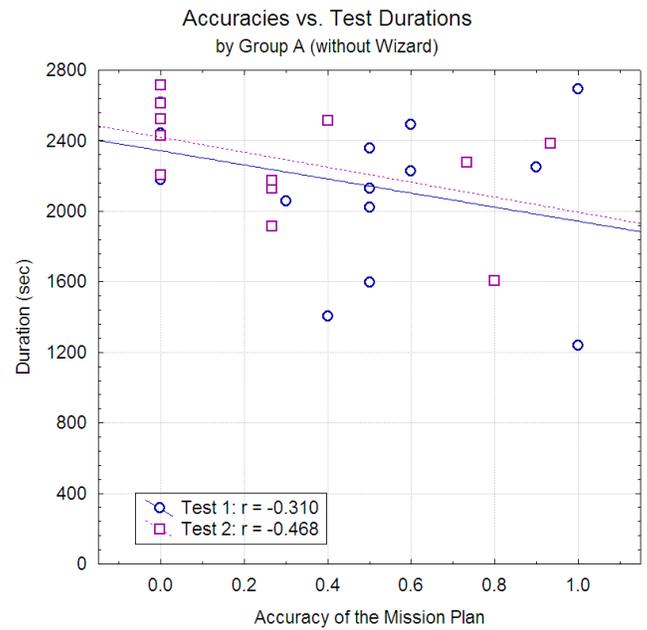


Figure 13: Scatter plot of the test durations against the accuracies for Group A (without the mission-planning wizard).

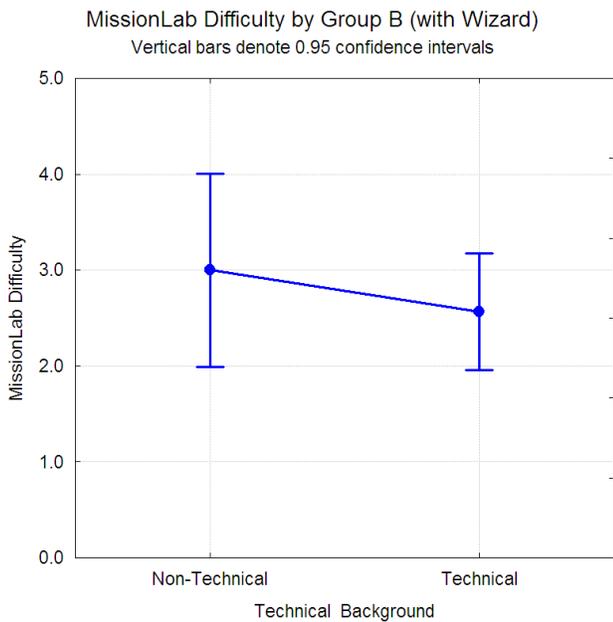


Figure 12: Comparison of the general *MissionLab* difficulty means between the Group B subjects (with the mission-planning wizard) with and without technical backgrounds. There was no significant difference between the two means ($p = 0.441$).

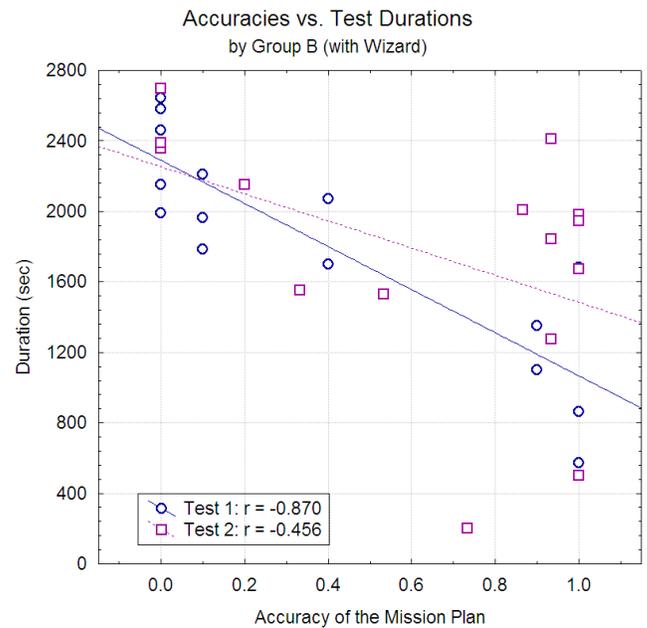


Figure 14: Scatter plot of the test durations against the accuracies for Group B (with the mission-planning wizard).

4) Accuracies vs. Test Durations

In order to analyze whether the quality of the mission plans was compromised by subjects rushing to finish the test, a linear correlation (Pearson's correlation coefficient r) between accuracy and test-duration was computed by *STATISTICA* for each group. The correlations are shown as scatter plots in Figure 13 for Group A (without the mission-planning wizard) and Figure 14 for Group B (with the mission-planning wizard). The r value ranges between -1 and 1 . If the value is larger than 0 , the relationship is direct, and if it is less than 0 , the relationship is inverse. As the absolute value of r approaches 1 , the relationship is stronger [7]. Interestingly, reverse relationships (i.e., the subjects who took less time tended to have higher accuracy) were observed for all the cases. The strength of the relationships ranged from weak ($r = -0.310$; Group A in Test 1) to strong ($r = -0.870$; Group B in Test 1). We therefore conclude that participants "hurrying to finish" did not compromise the quality of the mission plans, as we did not observe any direct relationships between those elements.

IV. CONCLUSIONS

With regard to the mission-planning wizard, the experiments demonstrated that it did indeed help reduce the amount of time required to plan a robot mission and increase the accuracy of the mission plan in Test 2. However, in Test 1, such advantages were not observed. This phenomenon is rather striking as the mission in Test 1 was simpler than the one in Test 2. A few speculative explanations can be made; 1) The mission-planning wizard works well in a complex mission, but not in a simple one; 2) During the Test 1 period, the subjects were still trying to figure out how to use the mission-planning wizard; or 3) Sampling size was still too small to obtain the appropriate statistical significance. In order to determine this, however, a different set of experiments would need to be conducted. Another advantage of the mission-planning wizard was that it compensated for a user's lack of a technical background, in terms of ease of use.

In this article, we have introduced guidelines for how to prepare usability experiments to evaluate robot mission planning systems. Since few robotics researchers assess the usability of their systems by conducting formal usability studies, these guidelines should provide value to the robotics community in and of themselves. We admit that, however, for some readers, the entire usability testing process (preparation, execution, and analyses) may appear tedious and costly. However, our experience suggests that once a usability experiment has been conducted, preparation of subsequent experiments can be completed more efficiently, as much of the procedures and documentation can be reused. Thus, we recommend that robotics researchers complete at least one formal usability experiment on their software as early as possible in order to establish a baseline for future studies.

ACKNOWLEDGEMENT

The authors would like to thank: Dr. Thomas Collins for reviewing various documents for the usability experiment; Amin Atrash, Michael Kaess, Brian Lee, Collin Loeb, and Eric Martinson for their support of the usability experiments; and Ron Snyder, Qing Lin, and Sridevi Salagrama for their contributions to *miCBR* and mission-planning wizard development.

APPENDIX

As examples, the protocols for Test 1 and Test 2 are included below. Details of other documents used in this experiment can be found in [12]. The results of significance testing are shown in Appendix C.

A. Protocol Used in Test 1

Test 1: Biohazard Mission (Subject Copy)

Mission:

Approach a building and conduct an indoor assessment for biohazard contamination.

Background:

It is 7:00 AM, and you are in a foreign country (somewhere in Asia). You are part of the USA military robotics unit sent by the United Nations to investigate whether the economic sanctions against this country can be lifted. The sanctions were originally imposed due to its possession of biological weapons. According to the CIA, the building that is approximately 100 meters away from where you are standing may possibly be a factory for building biological weapons.

Overall Task:

Your task is to create a robot mission plan where the robot approaches the building by following a series of waypoints. Once inside the building, it should proceed along the hallway and search every open room to check if biohazards are present. According to the CIA, the layout of this building is conventional; it has a linear corridor with rooms on both the left and right sides. However, the exact floor plan of this building is not known. The building is known to be vacant.

Note:

You have up to 45 minutes to finish this task.

You are done with this session if you either: (1) create the mission and you are satisfied with the simulation result, or (2) until the 45 minutes are over.

If you want to discard your work and start over, restart the MissionLab using Start Over button.

Detailed Instructions:

Your robot is currently at StartPlace shown in "factory.ovl" (See Figure 16). From this position, create a series of waypoints to bring the robot inside the building (labeled as Possible Factory for Biological Weapons). Once the robot is at the entrance door, it should proceed down the hallway. When it encounters a door, it should enter the room and conduct a biohazard survey. If the survey indicates that a biohazard is present, the robot should immediately alert the user. If the survey results do not indicate the presence of a biohazard, the robot should mark the room as visited, and continue the search, surveying additional rooms as needed. If the robot ever encounters a biohazard and alerts the user, its mission is completed and it can safely terminate the mission. If the robot reaches the end of the hallway, it should turn around and travel back down the hallway, searching for additional rooms it may have not checked.

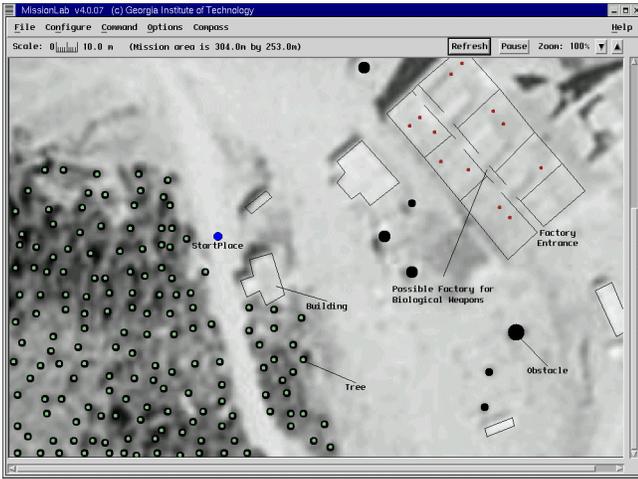


Figure 16: Overlay used in Test 1

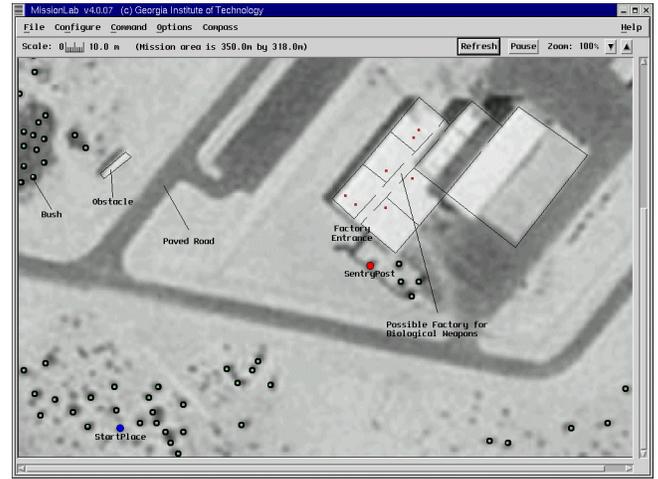


Figure 17: Overlay used in Test 2

B. Protocol Used in Test 2

Test 2: Biohazard Mission Part II (Subject Copy)

Mission:

Approach a building and conduct an indoor biohazard assessment. This time, use two robots.

Background:

It is 9:00 PM, and you are in a hostile territory (somewhere in Asia). This region of the country is controlled by a militant group who is known for its terrorist acts. The USA ground forces are expected to occupy this region within 5 hours. However, recent intelligence reports suggested that a building in this area might be a biological weapons factory. An indoor assessment of this building is crucial because unprepared ground troops should not enter if a biohazard is present.

Overall Task:

Your task is to create a two-robot mission in which the robots approach the building following a series of waypoints. Once inside the building, they should proceed along the hallway and search every room that has an open door for the presence of biohazards. According to the CIA, the layout of this building is conventional; it has a linear corridor with rooms on both the left and right sides. However, the exact floor plan of this building is not known. The sentry is very likely to be present in the building. (Page Break)

Note:

- You have up to 45 minutes to create this robot mission.
- You are done with this session if you either: (1) create the mission and you are satisfied with the simulation result, or (2) when the 45 minutes are over.
- If you want to discard your work and restart, Use the Start Over button in MissionLab.

Detailed Instructions:

Your robots are currently at StartPlace shown in the "hostile_territory.ovl". (See Figure 17) From this position, create a series of waypoints to bring the robots inside the building (labeled as Possible Factory for Biological Weapons). Once the robots are at the entrance door, they should proceed down the hallway. At any and every open door they encounter, one robot should enter the room and conduct a biohazard survey while the other robot remains in the hallway, near the doorway to that room, watching for a sentry to appear. If a sentry is detected, it should join the survey robot in the room and wait until the sentry leaves. If both robots are in the hallway when the sentry appears, they should both duck into the same nearby open room. If a biohazard is detected anywhere in the building, the robot should immediately alert the user, terminate the mission. As each room is checked and found to be clean, it should be marked as visited. The robots will continue their search until they reach the end of the hallway, when they should turn around and travel back down the hallway, checking any additional rooms they may have missed.

C. Significance Testing

Statistical Hypotheses:

- The mean of the Test 1 durations by Group A is significantly higher than the mean of the Test 1 durations by Group B.
- The mean of the Test 2 durations by Group A is significantly higher than the mean of the Test 2 durations by Group B.
- The mean of the Test 1 accuracies by Group B is significantly higher than the mean of the Test 1 accuracies by Group A.
- The mean of the Test 2 accuracies by Group B is significantly higher than the mean of the Test 2 accuracies by Group A.
- The mean of the general *MissionLab* difficulty by Group A is significantly higher than the mean of the general *MissionLab* difficulty by Group B.
- The mean of the general *MissionLab* difficulty expressed by the subjects without the technical background in Group A is significantly higher than the mean of the general *MissionLab* difficulty expressed by the subjects with the technical background in Group A.
- The mean of the general *MissionLab* difficulty expressed by the subjects without the technical background in Group B is significantly higher than the mean of the general *MissionLab* difficulty expressed by the subjects with the technical background in Group B.

ANOVA Results:

Hypo.	ANOVA Table						Validity of Hypo.
		SS	df	MS	F	p	
a	Group	898508	1	898508	2.998	0.095	Invalid p > 0.05
	Error	8092984	27	299740			
b	Group	1837883	1	1837883	6.131	0.020	Valid p < 0.05
	Error	7794327	26	299782			
c	Group	0.062	1	0.062	0.404	0.530	Invalid p > 0.05
	Error	4.126	27	0.153			
d	Group	0.849	1	0.849	5.881	0.023	Valid p < 0.05
	Error	3.751	26	0.144			
e	Group	0.436	1	0.436	0.619	0.438	Invalid p > 0.05
	Error	18.990	27	0.703			
f	Tech.	3.506	1	3.506	11.458	0.005	Valid p < 0.01
	Error	3.672	12	0.306			
g	Tech.	0.547	1	0.547	0.631	0.441	Invalid p > 0.05
	Error	11.261	13	0.866			

REFERENCES

- [1] Adams, J.A. *Human Management of a Hierarchical System for the Control of Multiple Mobile Robots*. Ph.D. Dissertation, University of Pennsylvania, 1995.
- [2] Ali, K.S. *Multiagent Telerobotics: Matching Systems to Tasks*. Ph.D. Dissertation, College of Computing, Georgia Institute of Technology, 1999.
- [3] Arkin, R.C. *Behavior-Based Robotics*. MIT Press, Cambridge, MA, 1998.
- [4] Arkin, R.C. and Balch, T. "AuRA: Principles and Practice in Review." *Journal of Experimental and Theoretical Artificial Intelligence*. Taylor and Francis, London, Vol. 9, No. 2, 1997, pp. 175-189.
- [5] Arkin, R.C., Collins, T.R., and Endo, Y. "Tactical Mobile Robot Mission Specification and Execution." *Proceedings for SPIE: Mobile Robots XIV*. The International Society for Optical Engineering, Bellingham, Vol. 3838, 1999, pp.150-163.
- [6] Buhler, C., Hoelper, R., Hoyer, H., and Humann, W. "Autonomous Robot Technology for Advanced Wheelchair and Robotic Aids for People with Disabilities." *Robotics and Autonomous Systems*. Elsevier Science Publishers, Amsterdam, Vol. 14, 1995, pp. 213-222.
- [7] Cohen, P.R. *Empirical Methods for Artificial Intelligence*. The MIT Press, Cambridge, MA, 1995.
- [8] Collins, T.R., Arkin, R.C., Cramer, M.J., and Endo, Y. "Field Results for Tactical Mobile Robot Missions." *Unmanned Systems 2000*. Association for Unmanned Vehicle Systems International, 2000.
- [9] Draper, J. "Human Factors in Telemanipulation: Perspective from the Oak Ridge National Laboratory Experience." *SPIE Telemanipulator Technology and Space Telerobotics*. The International Society for Optical Engineering, Bellingham, Vol. 2057, 1993, pp. 162-173.
- [10] Draper, J. "Teleoperators for Advanced Manufacturing: Applications and Human Factors Challenges." *International Journal of Human Factors in Manufacturing*. Wiley, New York, Vol. 5, No. 1, 1995, pp. 53-85.
- [11] Draper, J. and Blair, L. "Workload, Flow, and Telepresence during Teleoperation." *Proceedings: IEEE International Conference on Robotics and Automation*. IEEE Computer Society Press, Los Alamitos, Vol. 2, 1996, pp. 1030-1035.
- [12] Endo, Y., MacKenzie, D.C., and Arkin, R.C. *Usability Evaluation of High-Level User Assistance for Robot Mission Specification*. Georgia Tech Technical Report GIT-GOGSCI-2002/06, College of Computing, Georgia Institute of Technology, 2002.
- [13] DeRoos, B.G., Price, J.D., and Reidy, J.J. "Law Enforcement Robot Technology Assessment." *Proceedings of SPIE: Enabling Technologies for Law Enforcement and Security*. The International Society for Optical Engineering, Bellingham, Vol. 4232, 2001, pp. 409-420.
- [14] Georgia Tech College of Computing and Georgia Tech Research Institute. *Real-Time Cooperative Behavior for Tactical Mobile Robot Teams: Final Report A003*. Georgia Institute of Technology, 2001 [<http://www.cc.gatech.edu/ai/robot-lab/tmr/project.htm>].
- [15] Georgia Tech Mobile Robot Laboratory. *MissionLab: User Manual for MissionLab version 5.0*. Georgia Institute of Technology, 2002 [<http://www.cc.gatech.edu/ai/robot-lab/research/MissionLab/>].
- [16] Hackos, J.T. and Redish, J.C. *User and Task Analysis for Interface Design*. Wiley Computer Publishing, 1998.
- [17] Keates, S., Clarkson, J., and Robinson, P. "Designing a Usable Interface for an Interactive Robot." *ICORR '99: 6th International Conference on Rehabilitation Robotics*. Stanford, 1999, pp. 156-162.
- [18] MacKenzie, D.C. and Arkin, R.C. "Evaluating the Usability of Robot Programming Toolsets." *The International Journal of Robotics Research*. MIT Press, Cambridge, Vol. 17, Issue 4, 1998, pp. 381-401.
- [19] MacKenzie, D.C., Arkin, R.C. and Cameron, J.M. "Multiagent Mission Specification and Execution." *Autonomous Robots*. Kluwer Academic Publishers, Boston. Vol. 4, Issue 1, 1997, pp. 29-52.
- [20] Montemerlo, M., Pineau, J., Roy, N., Thrun, S., and Verma, V. "Experiences with a Mobile Robotic Elderly Guide for the Elderly." *National Conference on Artificial Intelligence*, AAAI, 2002.
- [21] Nguyen, H.G. and Bott, J.P. "Robotics for Law Enforcement: Application Beyond Explosive Ordinance Disposal." *Proceedings of SPIE: Enabling Technologies for Law Enforcement and Security*. The International Society for Optical Engineering, Bellingham, Vol. 4232, 2001, pp. 433-454.
- [22] Parsons, B., Warner, P., White, A., and Gill, R. "An Adaptive User Interface and Controller for Rehabilitation Robotic Arm." *ICAR '97: Proceedings of 8th International Conference on Advanced Robotics*. IEEE, Piscataway, 1997, pp. 919-923.
- [23] Rubin, J. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley and Sons, Inc., 1994.
- [24] Simpson, R., Poirot, R. and Baxter, M. "Evaluation of the Hephaestus Smart Wheelchair System." *Proceedings of the 6th International Conference on Rehabilitation Robotics*. Board of Trustees of Stanford University, Stanford, 1999, pp. 99-105.
- [25] Spyridakis, J.H. "Conducting Research in Technical Communication: the Application of True Experimental Designs." *Technical Communication*, Vol. 39, No. 4, 1992, pp. 607-624.
- [26] Virzi, R.A. "Refining the Test Phase of Usability Evaluation: How Man Subjects Is Enough?" *Human Factors*. The Human Factors Society, Inc., Vol. 34, No. 4, 1992, pp. 457-468.