In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*, NIST, Gaithersburg, MD, August 2002.

Evaluating the Performance of Assistive Robotic Systems

Holly A. Yanco
Computer Science Department
University of Massachusetts Lowell
One University Avenue
Lowell, MA 01854
holly@cs.uml.edu

ABSTRACT

When designing robotic systems for the disabled, it is necessary to demonstrate that the systems are safe. Beyond safety, it must also be shown that the equipment improves the quality of life for its user. This paper discusses methods for testing assistive robotic systems to assure safety, usability and usefulness. An example of a user test designed for a robotic wheelchair is presented and discussed.

KEYWORDS: *Assistive technology, assistive robotics, robotic wheelchair, user tests*

1 INTRODUCTION

Assistive technology enables people to do things that would be impossible without the technology. Assistive technology can range from smart homes for the elderly to robotic wheelchairs to voice control software for a computer. Robotic workstations can provide people with vocational assistance [Dallaway et al 1995, Kazi et al 1998, Wagner et al 1999]. Robotic walkers can be used to allow elderly people with decreased vision to walk around their nursing home [Lacey 1999].

In this paper, we discuss the testing process for assistive mobile robots which either carry or lead their users, so that the user and the robot must travel together during the use of the assistive system. We will call these devices assistive navigation systems. We will use Wheelesley, a robotic wheelchair system developed for indoor and outdoor use, in our discussions to illustrate the evaluation methods [Yanco 2000].

The target community for a robotic wheelchair system consists of people who are unable to drive a powered wheelchair using a standard joystick. This group includes people with cerebral palsy, stroke patients who omit stimuli from one side, and quadriplegics. The users vary in ability and access methods[1] used to drive the wheelchair. Some people can move a joystick, but are unable to make fine movement corrections using the joystick. Other people are able to click one or more switches using their head or other body part. Some potential users are unable to control a powered wheelchair with any of the available access devices and must rely upon a caregiver to move them throughout the world. A robotic wheelchair will enable this population to better self-navigate through the world, increasing independence.

Human-robot interaction must be considered when designing assistive travel systems. Designing a poor interface will result in an unusable system. Robotic wheelchairs must be able to connect to a variety of commonly used access methods. User tests must utilize the access methods to be used by the target population, even when testing able-bodied subjects. With a target population lacking the fine motor control necessary to move a joystick, user tests with able-bodied subjects using a joystick can not be extrapolated to the intended users.

2 EVALUATORS FOR SYSTEMS

When designing and evaluating assistive navigation systems, there are three groups of people that should be involved: providers, able-bodied test subjects and people in the target population of the system.

Providers are the people who prescribe and deliver systems to a user. In the case of wheelchairs, physical therapists adapt a wheelchair to its user by creating custom cushions, determining the proper access method, and adjusting settings such as speed controls. Physical therapists also work with wheelchair users to teach them how to use the system. Since they are very involved with providing care to users, it is important to involve these providers from the early stages of development through the final product testing. People who work individually and daily with users will have an understanding of the needs of the population.

Systems should be tested first on able-bodied subjects. Many members of the target population for robotic wheelchairs are non-verbal, making it difficult to do user tests since it is important to be able to tell if the user feels comfortable when testing the system. Walkers for the elderly infirm who have limited vision should first be tested on people who can see. The move to a target user should be made only after the safety and reliability of the system has been repeatedly demonstrated through able-bodied user tests.

[1] An access method is a means for controlling a powered wheelchair, such as a joystick or a sip-and-puff system.

3 EVALUATION OF SYSTEMS

The performance of an assistive navigation system must be measured through user tests. Tests should range from preliminary demonstrations of safety and usability to long term use by one or two subjects from the target population.

Before testing the system on users from the target population, tests on a large number of able-bodied subjects should be undertaken. These tests usually involve a test course, where subjects must traverse the course using assisted navigation and unassisted navigation for comparison. Metrics to be collected include time required to travel the course, number of safety violations which may range from scrapes to bumper hits to more serious failures, and the amount of effort required to drive the course.

A system's performance should be measured by conducting user tests that compare the performance of the robotic system against a non-robotic solution. For example, a robotic wheelchair can be designed to allow it to be controlled with no sensor mediation (manual control) or with sensor-mediated (robotic) control. By designing a system this way, the same user interface and access methods can be used to compare user performance with assisted control and with unassisted control. The human-robot interaction must be duplicated even when the robotic control is not used for navigation. It would be impossible to directly compare a traditional powered wheelchair driven by a joystick to a robotic system driven with an access method such as single switch scanning. This difference is even more problematic due to the fact that the target community does not have the fine motor control necessary to drive a powered wheelchair using a joystick. Care must be taken to design user tests that change only a small number of variables (ideally one).

Tests involving the target users should focus on long term use instead of runs on a test course. Before these long term tests of a small group of target users, researchers should consider undertaking long term tests of a few able-bodied subjects so that the test methods may be fully evaluated before proceeding to test subjects with reduced mobility.

4 ABLE-BODIED USER TEST DESIGN

To illustrate the design of able-bodied user tests, this section discusses indoor user tests designed for and performed with the Wheelesley robotic wheelchair system [Yanco and Gips 1998]. An experiment to test the performance of able-bodied subjects under robotic assisted control and under standard manual control was designed to determine if robotic assistance improved driving performance using single switch scanning as an access method. Single switch scanning is the access method of last resort for powered wheelchairs, primarily because drift is a significant problem. To correct a drift to the left or the right, the user must stop going forward, wait for the scanning device to get to the arrow for the

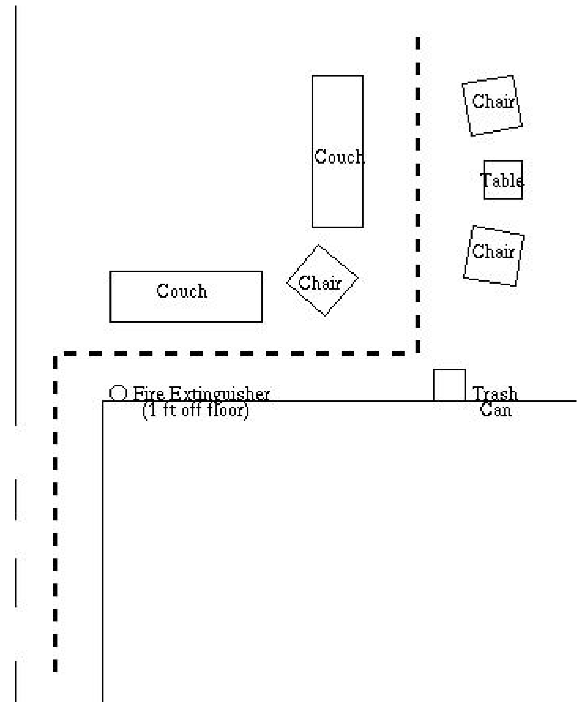

Figure 1: The indoor test course. A test run consisted of an up and back traversal of this course.

direction of choice, click to turn the chair, stop turning, wait to scan to forward and then click to move forward again.

Fourteen able-bodied subjects (7 men and 7 women), ranging in age from 18 to 43, were tested. All subjects were familiar with using computers and none had driven the wheelchair before.

At the beginning of a session, the subject was shown the wheelchair. Sensors that are used in robotic assisted control were pointed out and explained briefly. Safety measures, such as the power button, were discussed. Then the two driving methods were explained to the subject. After this introduction, the subject was seated in the wheelchair and the user interface was connected to the wheelchair. The single switch scanning interface was explained to the subject, who then practiced using the interface with the motors turned off.

Once the subject was comfortable with the interface, the session entered a practice phase in which the subject first tried robotic assisted control and then standard manual control. The subject practiced both methods until he expressed an understanding of each control method; subjects usually spent about two minutes trying each method. All practice was done off of the test course, so that the subject was not able to learn anything that would assist him during the test phase.

The course (shown in Figure 1) was designed to include obstacles (several couches and chairs, a fire extinguisher mounted to the wall 30 cm (11.8 inches) above the ground, a trash can, and a table) and turns to the left and to the right. The course is 20 meters (65.7 feet) long. Three doors in the

	Manual				Robotic			
	First Run		Second Run		First run		Second run	
Number of clicks	90.2	(16.3)	77.1	(9.8)	25.6	(4.9)	22.0	(3.3)
Scanning time (sec)	93.6	(20.3)	81.1	(13.0)	30.9	(8.3)	25.2	(8.6)
Moving time (sec)	311.6	(36.4)	316.6	(36.2)	268.2	(21.5)	277.1	(28.4)
Total time (sec)	405.1	(42.1)	397.7	(43.7)	299.1	(18.4)	302.3	(32.5)

Table 1: Results of the indoor user tests: the number of clicks, amount of time spent scanning for commands, amount of time moving and total time to complete the course. The first number for each method is the mean and the number in parentheses is the standard deviation.

hallway on the course could be open or closed, determined by the office occupants.

The test phase consisted of four up-and-back traversals of the test course, alternating between the two control methods. Half of the subjects started with robotic assisted control and the other half started with standard manual control. Each up-and-back traversal consists of two parts: running the course from the couch area to the hallway and then the return trip. The turn in the middle of the course is not counted as part of the run, as turning completely around in the middle of the hallway is not a normal driving occurrence. The total session time for each subject was approximately 45 minutes.

There were four experimental performance measures collected by the computer that was running the user interface: (1) the number of clicks required to navigate the course, (2) the amount of time spent scanning to get to the necessary commands, (3) the amount of time spent moving or executing the given commands, and (4) the total amount of time spent on the course (scanning time plus moving time). The researcher only recorded the number of scrapes made by the chair. At the completion of the test, the user was asked to rank standard manual control and robotic assisted control on a scale from 1 (worst) to 10 (best).

Data for each experimental measure was analyzed using an ANOVA test. The differences between robotic control and manual control were highly significant with $p<.0001$ for all measures. On average, robotic control saved 60 clicks over manual control, which is a 71% improvement. Total time for robotic assisted control was 101 seconds shorter than manual control on average, which is a 25% improvement.

The only performance measure not collected on the computer was the count of the number of scrapes. A scrape was recorded when the chair brushed along a wall or piece of furniture. Bumps with the bumper were also counted as scrapes. No subject hit a wall or an obstacle with great force. The average number of scrapes per run under manual control is 0.25. The average number of scrapes under robotic control is 0.18. These numbers are not significantly different.

Finally, the subjects were asked to evaluate the two driving methods by giving a score from 1 (worst) to 10 (best). The average score for standard manual control was 3.5. The average score for robotic assisted control was 8.7. These scores are highly significant with $p<.0001$. No test subject preferred manual control over robotic control.

Subjects drove more efficiently and preferred to drive with robotic assisted control. Robotic control automatically adjusts for drift where manual control does not. When traveling down a long hallway under robotic control, a user can click on forward at the beginning of the corridor and does not need to do anything more until he wishes to stop or turn. Under manual control, the user must make many adjustments to compensate for drift.

The total time taken on a test run is a sum of the scanning time and the command execution time. Both scanning time and execution time improved from manual control to robotic control. As would be expected, if fewer clicks are issued, the scanning time required is shorter. Estimating that forward is clicked 50% of the time, left is clicked 25% of the time and right is clicked 25% of the time[2], with a scan time of one second and an estimated reaction time of one half second in our able-bodied subjects, each click would require an average of 1.25 seconds. As Table 1 shows, the scanning time is approximately 1.25 times the number of clicks.

Each user executed two trials for each control method. The differences between the two trials were significant for clicks ($p=.003$) and for time spent scanning ($p=.015$), indicating that the subjects were improving due to learning.[3] As the user became more comfortable with the system, he was able to judge more effectively when it was necessary to make adjustments to the current course.

A single subject ran the course 10 times in manual mode to determine how learning could affect the number of clicks and scanning time. The subject was this researcher; a naive user is not required to test for optimal performance. Over the 10 runs, the average number of clicks in a test run was 71.4 with a standard deviation of 9.5. Over the last 5 runs, the average number of clicks was 68 with a standard deviation of 4. Scanning time averages 73.7 seconds (standard deviation

[2] Empirically, backwards commands are issued very infrequently.

[3] There was no significant effect of learning on moving time and total time between trials; since the speed is held constant throughout the experiment, the user can not significantly reduce the amount of time required to travel the course between trials of the same control method.

12.0) over all 10 runs and 68 seconds (standard deviation 5.3) over the last 5 runs. Optimal performance for this course in manual mode will not approach the average performance in robotic mode.

5 COMPARABLE ABLE-BODIED TESTS

The NavChair system was tested in an indoor environment using voice control as the access method [Simpson and Levine 1997]. Six able bodied subjects navigated through three different scenarios (room traversal, door passage and wall following), four times with navigation assistance and then four times with no navigation assistance. For each scenario, it took longer to navigate using navigation assistance, primarily because the chair slows down as it gets closer to obstacles. However, no collisions occurred with navigation assistance, while there were occasional collisions with no navigation assistance. Test subjects preferred driving with navigation assistance.

The VAHM Project was also tested in an indoor environment using single switch scanning as the access method [Bourhis and Pino 1996]. Four able bodied users familiar with computers drove the wheelchair through a course simulating a kitchen and living room environment in manual mode and in an assisted mode which provided obstacle avoidance. Assisted mode resulted in a 13% improvement in the number of actions required on the interface screen (which we called the number of clicks required). Execution time also improved an average of 7.7%. The tests were executed at three scanning rates: 0.8, 2.5 and 4.5 seconds. There was a 2.5% improvement in execution time for the 0.8 second rate,[4] a 7.7% improvement for the 2.5 second rate, and a 13% improvement for the 4.5 second rate. One would expect to see a more dramatic improvement in total execution time for longer scanning times since fewer clicks result in a greater time savings.

Indoor user tests of Tin Man II used a joystick and buttons for the access method [Miller and Slack 1995]. The test course was 50 meters long and included a hallway, a doorway and two rooms. Subjects were told to attempt to minimize their joystick movements in both the manual and assisted tests. Manual mode required 50% more joystick moves than the obstacle avoidance mode. The time required to traverse the course was less than 10% longer in assisted mode than in manual mode.

6 CONCLUSIONS

User tests are required for assistive navigation systems. These systems are designed for people to ride; it is important that a great deal of testing occurs from the early design stages to the end. Able-bodied test subjects should be utilized for all testing until the system is demonstrated to be safe, reliable and useful. Only after a long period of testing on able-bodied subjects should the testing proceed to the target population.

When designing assistive technology, providers should be asked for comments on the system from the initial design stages through final testing of the product. The inclusion of providers will result in designs that better reflect the needs and desires of the target population. Providers can also facilitate safe testing of the target population.

Performance metrics for assistive navigation systems are necessary to demonstrate the usefulness and usability of the system. Assistive navigation systems are medical devices; as such, they should be held to the same strict testing guidelines mandated by the FDA, even in the initial design phases.

[4] This is the closest to our 1 second scanning rate.

ACKNOWLEDGMENTS

The Wheelesley user tests were performed while the author was a graduate student in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. This research was funded by the Office of Naval Research under contract number N00014-95-1-0600 and the National Science Foundation under grant number CDA-9505200.

REFERENCES

[Bourhis and Pino 1996] G. Bourhis and P. Pino. "Mobile robotics and mobility assistance for people with motor impairments: rational justification for the VAHM project." *IEEE Transactions on Rehabilitation Engineering*, 4(1): 7 – 12, March 1996.

[Dallaway et al. 1995] John L. Dallaway, Robin D. Jackson, and Paul H. A. Timmers. "Rehabilitation robotics in Europe." *IEEE Transactions on Rehabilitation Engineering*, 3(1):35 – 45, March 1995.

[Kazi et al. 1998] Zunaid Kazi, Shoupu Chen, Matthew Beitler, Daniel Chester, and Richard Foulds. "Speech and gesture mediated intelligent teleoperation." In *Lecture Notes in Artificial Intelligence: Assistive Technology and Artificial Intelligence*, edited by Vibhu Mittal, Holly A. Yanco, John Aronis, and Richard C. Simpson, Springer-Verlag, pp. 194 – 210, 1998.

[Lacey 1999] Gerard Lacey. *Personal Adaptive Mobility Aid for the Frail Visually Impaired*. Ph.D. Thesis, University of Dublin, Trinity College, April 1999.

[Miller and Slack 1995] David P. Miller and Marc G. Slack. "Design and testing of a low-cost robotic wheelchair prototype." *Autonomous Robots*, 2:77 – 88, 1995.

[Simpson and Levine 1997] Richard C. Simpson and Simon P. Levine. "Development and evaluation of voice control for a smart wheelchair." In *Proceedings of the Rehabilitation Engineering Society of North America 1997 Annual Conference*, pp. 417 – 419, RESNA Press, 1997.

[Wagner et al. 1999] J.J. Wagner, H.F.M Van der Loos, N. Smaby, K. Chang, and C. Burgar. "ProVAR assistive robot interface." In *Proceedings of ICORR '99: International Conference on Rehabilitation Robotics*, pp. 250 – 254, 1999.

[Yanco and Gips 1998] Holly A. Yanco and James Gips. "Driver performance using single switch scanning with a powered wheelchair: robotic assisted control versus traditional control." In *RESNA-98*, pp 298 – 300, 1998.

[Yanco 2000] Holly A. Yanco. *Shared User-Computer Control of a Robotic Wheelchair System*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, September 2000.