Metrics for Human Driving of Multiple Robots

Dan R. Olsen Jr, *Member IEEE Computer Society*, Stephen Bart Wood, and Jonathan Turner

Abstract—**A goal of human-robot interaction is to expand the capability of human beings so that they can control multiple robots simultaneously. This provides leverage for human attention beyond single robot interaction. The number of such robots that can be operated is called the fan-out of a human-robot team. We define the concepts of neglect-tolerance as a measure of autonomy and interaction-effort as a measure of required human attention. A model for fan-out is presented that predicts a human's ability to control multiple robots. We validate this model using robot simulations and user studies. Using the validated model we demonstrate enhanced task performance on physical robots as predicted by the model.**

Index Terms—**human-robot interaction, multiple robots, fan-out.**

I. INTRODUCTION

For the foreseeable future, robots will not be fully autonomous but will be guided and directed by humans. The goal in our lab is to find ways to expand human ability to control robots. In particular we want ways to leverage human attention to accomplish tasks more effectively. One approach is to allow a human to drive many robots simultaneously. In pursuing this goal we need metrics to evaluate the effectiveness of a human-robot team. Is the interface responsible for the performance or the robot's autonomic capabilities. The difficulty with such evaluations is that the autonomy of the robot gets intertwined with the quality of the user interface. We want to tease these apart so that we can clearly identify where the improvements or difficulties are coming from so that designers and researchers can more effectively focus their efforts.

In this paper we look at a particular aspect of HRI, which is the ability for an individual to control multiple robots simultaneously. We refer to this as the *fan-out* of a human-robot team. The fan-out of a human-robot system is the average number of robots being using simultaneously. We hypothesize that the following *fan-out equation* holds,

$$FO=NT/IT$$

where
   FO=fan-out or the number of robots a human can control simultaneously,
   NT=neglect tolerance or the time that a robot can operate without human intervention,
   IT=interaction time or the time that it takes for a human to interact with a given robot.

Suppose that a robot can, on the average, operate for 10 seconds without human intervention and it takes only 2 seconds for the user to decide what needs to be done and give the robot new instructions. This would mean that the user would have 8 out of every 10 seconds available to devote attention to other work. In the case of a multi-robot team, this attention could be devoted to other robots. If each robot takes 2 seconds of attention and can then run on its own for 8 more seconds, then a user should be able to operate 5 robots simultaneously. This is the rationale for our fan-out model of multi-robot interaction.

In this paper we also describe how this model can be used to compute metrics of the effectiveness of the human-robot interface. We describe how we have validated the model using a simulated robot world and lastly we describe how the model was applied to physical robots being driven by novices using two different third-person "view from above" interfaces.

II. PRIOR WORK

Others have done work on human-robot interaction. Sheridan has outlined 5 levels of robot control by users [14]. The levels range from teleoperation, where the user is directly engaged with the actuators of the robot, through various levels of computer intervention between the user, the sensors and the actuators, to full autonomy with users merely setting very high-level goals. Fong and Thorpe [8, 9] demonstrated collaborative control where the human and the robot share the initiative, with the robot seeking guidance when needed. These with a variety of other approaches are characterized by their system architecture. Although human-robot interfaces are provided, there is little study of the nature of that interface nor on how to evaluate the quality of the interface.

There have been a number of proposals for new modalities for controlling robots including haptics, gestures and PDAS [7]. Others have looked at the visualization and context memory problems that arise when driving robots. The Egosphere is one such solution [6].

There is also a great deal of work on using multiple robots on a task. There are fully autonomous swarming approaches such as in Bruemmer, et al [3]. These have very little human intervention because the desired task is preprogrammed. Other autonomous robot teams have done janitorial tasks, box pushing and earth moving [12, 13]. All of these teams have used very little human intervention. Other multi-robot systems have robots operating in formations [2, 4, 16] or according to predefined deployment behaviors [15]. These approaches allow users to direct the work of a number of robots

simultaneously. Fong et. al. [10] point out the problems with dividing human attention among multiple robots and propose a collaborative control model for driving. In essence their proposals increase the neglect and activity time of the robots to achieve higher fan-out. Others have used a "select and command" model for controlling multiple robots [11].

However, none of these have been carefully evaluated as to the advantages or decrease in effort afforded by the various user interface designs. In most cases the control architecture is intertwined with the human-robot interface making it hard to distinguish which part of the solution is contributing to progress. In this paper we describe a model for isolating and measuring the human-robot interface for teams of robots.

III. A SIMPLE WORLD

To clarify the concepts behind our metrics we pose the simple robot world shown in figure 1. In this world, the user gives the robot a distance and a direction. The neglect tolerance of the robot is the amount of time that the user can ignore the robot and pay attention to others. From the diagram it would seem that this time would be the average time required to travel the distance of each leg of the planned course. However, the robot may be only partially capable of carrying out the user's instructions. Faulty odometry or lack of wheel traction may cause the robot to head into the trees and stop prematurely or may allow the robot to go beyond the end point and into the small cul-de-sac beyond. If the robot has very poor "crash detection" the user may give it much shorter, more cautious goals to prevent catastrophes. Thus the "trustability" of the robot also factors into to how long it can actually be neglected.

In most cases, increasing the intelligence of the robot can increase its neglect time. For example, adding simple sonar sensors to the robot to give it primitive obstacle avoidance would increase the user's ability to trust the robot (it is not as likely to hit a tree) and would allow the robot to proceed through more complex paths without human intervention.

Obviously neglect tolerance is a simplification of many interrelated factors. However, it does provide us with an overall measure of the autonomy level of the robot in a given task situation.
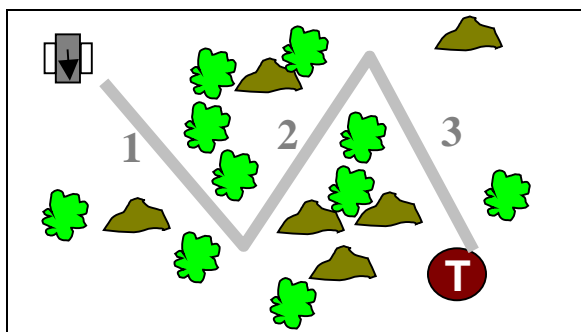

Fig 1. A simple robot world where the robot can be given a direction and a distance. The human must intervene at each leg of the journey or when premature failures occur.

A second factor in our fan-out model is the interaction time. This is also a simplification of many interrelated issues. The interaction time is the time it takes for a human to interact with a robot in order to give a single specific instruction. There are many components to interaction time most of which cannot be directly measured. The simplest measure of interaction time would be the amount of time required to express a new speed and direction for the robot. However, this is only part of the interaction time. If the robot is not trusted, then the user must expend time monitoring the status of the robot to make certain nothing bad happens. If the robot has no path planning ability then the user must spend more time planning feasible paths through the world and must make more interventions to accomplish the goal. As we introduce multiple robots to the user there is the thinking time required to switch contexts between robots. Most of the planning, monitoring and context acquisition time cannot be directly measured. This "unmeasurable" effort is a key challenge in developing our desired metrics.

IV. MEASURING HUMAN ROBOT INTERACTION

Given the definition of our model, we now look at how to use the model to measure the quality of the human robot interface.

A. Measuring Neglect Tolerance

The first step is to measure the neglect tolerance. Our first approach to measuring neglect tolerance was to randomly place a robot in a world and randomly choose a goal for the robot. We then measure the average time the robot can run before it stops making progress toward its goals. This turned out to be a very ineffective technique for measuring neglect tolerance. The first problem is that with real physical robots this becomes very time consuming. Secondly this approach does not take into account the trustablity of the robot. Our experiments have shown that this level of trust is very important to neglect tolerance.

The measure that we have found most effective in measuring neglect tolerance is activity time. Activity time is the time between when a the user gives a command to the robot and one of the following occurs:
  - the user gives a new command,
  - the robot stops making progress towards the goal,
  - or the robot reaches the goal.
We can give a user a team of robots and a task and then we can measure the average activity time of all of the robots. This will give us a good overall measure of the autonomy of those robots.

B. Measuring Fan-out

The most intuitively direct approach to measuring fan-out is the effectiveness plateau method. In this method we give a

user a task and a number of robots. We then measure the user's performance (time to complete or number of targets identified). We then increase the number of robots with more experiments until performance does not improve. This plateau in performance defines the maximum number of robots that can be effectively used. This approach is intuitively very attractive, but infeasible in practice. The problem is that to get statistically significant data one must make many runs with many users with many different numbers of robots.
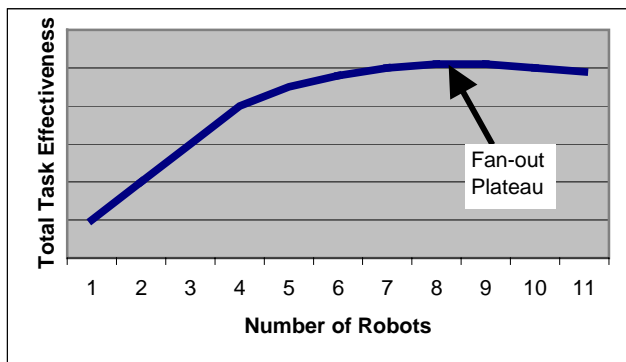

Fig 2. Measuring Fan-Out by repeated experiments increasing the number of robots until the performance does not increase with the addition of new robots.

We have found a more practical measure of fan-out to measure the number of robots making progress towards a goal at any point in time. We then average this over many time samples within a given experiment. For this measure to be used we need a predicate for "progress". A simple predicate is that the robot is moving and therefore making progress. However, if the robot is caught in some cycle it may move for a long time without making progress. Another predicate might be "getting closer to a goal." However, in some complex environments, the only feasible path to a goal may involve moving away to get around an obstacle. A fallback position is to manually code videotapes of the experiments to allow humans to judge "progress." Assuming we have a "making progress" metric for our task we can then give the user many robots and measure the average number of active robots at each time sample.

C. Measuring Interaction Effort

Measurements of neglect time and fan-out are useful in their own right, but what we really want is a measure of interaction effort. We do this by solving the fan out equation for interaction time (IT) giving

$$IT=AT/FO$$

However, because we are approximating neglect with activity time and approximating fan-out with average number of active robots, we do not believe that this equation will produce accurate interaction times. However, we do believe that this will produce a metric that can be used to compare two interfaces. We replace interaction time with a unitless measure called interaction effort (IE).

$$IE=AT/FO$$

For a given human-robots team we can measure average activity time, average number of active robots and then compute interaction effort. We can then change the user interface, perform experiments to give us a new interaction effort and measure any improvement. We thus have a technique for measuring the improvement in a human-robot interface.

D. Validating the Model

An unsettling feature of computing interaction effort from activity time and fan out is that it is an indirect measure computed from a model that seems rational, but is not necessarily valid. By computing rather than measuring one of the unknowns in the equation we loose our ability to verify its validity. Our fundamental problem is that interaction effort is, to a large degree, unmeasurable. Much of the effort goes on inside of the users head. The overt activities of operating the interface are generally only a minor part of the effort of controlling a team of robots. Before we can use this metric, we need a way to validate the model.

To illustrate our validation approach, suppose one built a set of robots and applied them to a task. Experiments are performed to measure AT and FO from which IE is computed. Reviewing the performance of the robots, the designers improve the obstacle avoidance of the robots, but do not change the human-robot interface. In this situation AT will go up because the robots can run more independently. FO will also go up because the user can drive more robots. If IE is really a valid measure of the human-robots interface then it should not change if the interface did not change.

V. ROBOT SIMULATIONS

To validate our fan out model we chose robot simulations rather than real robots. This allows us to more carefully control the experiment. The use of real robots will introduce many factors of robot reliability and terrain/task irregularity. Real robots would be better for actual measurements, but not as effective at validating the model due to numerous uncontrollable variables.

There is one way in which the real world differs sharply from our simulated world. In the real word, robots crash into obstacles, fall into holes, and run into each other. Safety is a real issue and lack of safety reduces the user's trust. As discussed earlier reduced trust leads to reduced activity times. In our simulations, robots never crash or fail therefore trust is higher than reality. However, we believe that this will be reflected in different activity times and should not affect the validity of the fan-out equation.

A. The task

For our fan-out experiments we chose a maze-searching task. We built a random maze generator that can automatically generate tasks of a given complexity. We defined task complexity as the dimensions of the maze, density of obstacles and number of targets. Using our random maze generator we were able to create a variety of tasks of a given complexity.

After random placement of obstacles and targets the maze was automatically checked to make certain that all targets were reachable. Our measure of task effectiveness was the time required for all targets to be touched by a robot.

To validate our fan-out model we created uniform tasks with a uniform user interface. What varies is the intelligence of the robots and their neglect time. If the fan-out equation holds, task times, neglect times and fan-out numbers will vary with the intelligence of the robots, but the interaction time should remain constant because the user interface is constant.

Figure 3 shows our maze running application and its interface. The users control the robots by dragging a goal location out in front of each robot. The robots attempt to reach the goal
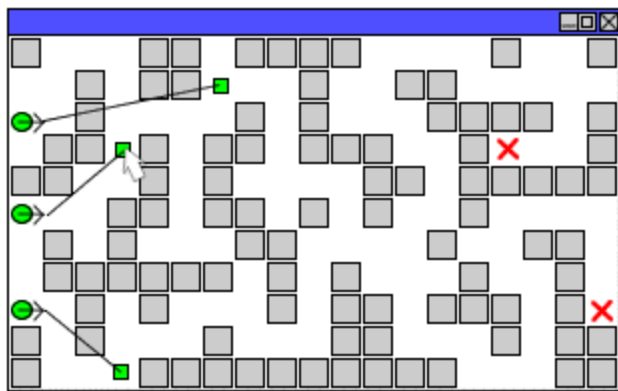

Fig 3. Driving robots(circles) through a maze to seek targets (Xx). User control is by dragging the square associated with each robot to set a short term goal for the robot.

In our initial experiments, users would visually solve the maze problem and then "micro-manage" the robots to get to the targets. For most of our tests we used the world shown in figure 4 where any area that has not been "seen" by a robot is obscured. This makes the task more of a robot control and space coverage task than a visual maze-solving task.
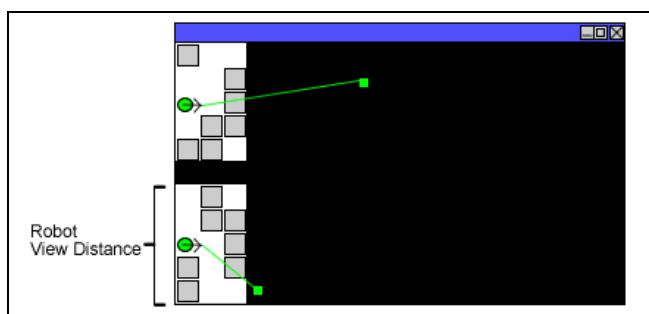

Fig 4. Obscured maze world. Only areas "seen" by a robot are shown.

B. Three types of robots

To test our fan-out theory of constant IE for a given user-interface we developed three types of simulated robots. The first type (*simple*) heads directly towards its current goal until

it reaches the goal or runs into an obstacle. This is a relatively simple robot with little intelligence.

The second type (*bounce*) bounces off obstacles and attempts to get closer to the goal even if there is no direct path. It never backs up and thus gets trapped in cul-de-sacs. The bouncing technique solves many simple obstacle avoidance problems but none that require any global knowledge. This robot stops whenever it cannot find a local movement that would get it closer to the goal than its current position.

The third type of robot (*plan*) has a "sensor radius". It assumes that the robot can "see" all obstacles within the sensor radius. It then uses a shortest path algorithm to plan a way to reach the point on its sensor perimeter that was closest to the goal. This planning is performed after every movement. This robot stops whenever its current position is closer to the goal than any reachable point in its sensor perimeter. This robot can avoid local dead-ends, but not larger structures where the problems are larger than its sensor radius.

C. User Tests on the Simulation

In attempting to perform fan-out tests we quickly learned that subject motivation has a lot to do with fan-out measurements. Simultaneous control of multiple robots requires more intense mental effort than controlling one. Unmotivated subjects tend to resist making the effort. In a real search and rescue scenario, where lives are at stake, such motivation will be much less of an issue. To overcome the motivation problems, we held a series of races with cash prizes for the first, second and third place winners in each race. For the most part this resolved the motivation issues.

In our first race there were 8 participants all running 8 races using the obscured worlds. The density of obstacles was 35% with 18 robots available and 10 targets to find. We ran 2 races with simple robots and 3 races each for the bounce and plan robots for a total of 64 trial runs. The measured fan-out and activity time along with the computed interaction time is shown in figure 6. Analysis of variance shows that there is no statistical difference in the interaction times across the three robot types. This validates the fan-out equation.

Robot Type	Mean Fan-out	Mean Activity Time	Computed Interaction Effort
Simple	1.46	4.36	3.06
Bounce	2.94	7.82	2.77
Plan	5.11	14.42	2.88

Fig 5. Measures of robot performance with three types of robots and the same user interface with 35% obstacles in the maze. Fan-out is the average number of robots active at any one time. Activity time is the average amount of time that a robot was running without stopping or human intervention.

Because our maze worlds are automatically generated we can control the complexity of the problem that the human-robot team must solve. We held a second competition with 8 subjects and 8 races where the obstacle density was reduced from 35% to 22%. With a less complex world the activity time

should go up. As the model predicted, the activity time rose as robots became more independent due to the reduced complexity of the world. However, fan-out also rose. The computed interaction effort is statistically unchanged from the more complex mazes. This is consistent with the fact that the interface was unchanged. Again the model is validated.

We ran a third competition where the world complexity was the same, but the interface was changed to a lower resolution. The lower resolution interface required the users to scroll and thus was more cumbersome. The interaction effort measurements were sharply higher as the model would predict.

VI. ACTUAL ROBOT EXPERIMENTS

To test our model on actual robots we created an environment where 4 small robots are placed in a maze with target locations for each robot, as shown in figure 6. The goal of the user is to get every robot to its appropriate target.
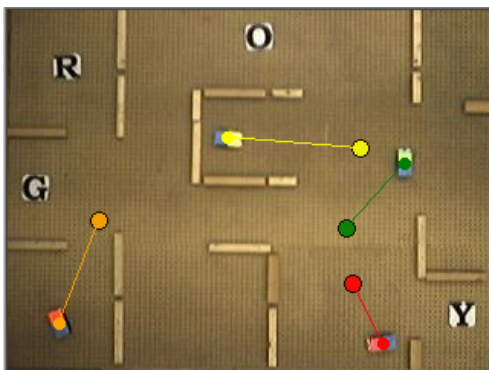

Fig 6. Robot maze problem. User is controlling robots to reach their corresponding goal. Control is by dragging a short term target out in front of each robot.

For this problem we had two control paradigms. The first was simple directional controls for each robot. These were represented by circles for each color of robot that would set its direction or stop. Our robots had no speed control. In this paradigm the user could control the direction of a robot and tell it to go. The users had to monitor the robots to make certain that they did not run into each other or the walls of the maze. The second control paradigm is shown in figure 6 where the user can drag out short-range goals for each robot. A camera-based tracking system would then monitor each robot's location and direct it towards its goal. The camera system was not capable of avoiding obstacles or other robots. The user had to do those tasks by setting a succession of achievable short-range goals.

We had 18 users make test runs with both types of control. We counter balanced the order to eliminate learning effects between control types. We collected the fan-out, activity time and total task time for all of the runs. The statistics are shown in figure 7. With direction driving, a user could barely keep one robot operating at a time. With the camera controlled goal-seeking control, the fan-out rose to 2.2 robots operating

at once. The activity time was also sharply higher because robots could operate without intervention for longer periods of time. Because of the parallelism that the camera-control offers, the total task time was almost ¼ of the time required by circle driving. Because the user was switching attention among multiple robots, the interaction effort to control the camera-controlled robots was slightly higher. In this case the fan-out model provides us with informative information about where the improvements are coming from. This is exactly what we wanted from our metrics.

	Direction	Camera
Average fan-out	0.93	2.20
Average activity time	1.89	5.27
Average task time	223.26	60.66
Computed interaction effort	2.02	2.39

Fig 7. Measurements of small robots in a maze using simple directional controls (Circle driving) and goal-based camera-managed controls (Camera).

VII. CONCLUSIONS

We can measure the parallelism in a human-robots team by measuring the fan-out, or the average number of robots making progress at any one time. We also measure the activity time or average time that the robot makes progress without human intervention. Using these two we compute interaction effort, which we can use to compare two systems to see if the user interface has improved. Our simulations and our robot experiments seem to bear out the hypothesis that the fan-out law holds and can pull out the human/robot contributions to task effectiveness.

In working with our robots and the model we understand that these metrics sweep many diverse issues under the rug. However, they do allow us gage overall improvements and thus give us good system-wide measures of progress.

REFERENCES

[1] Asimov, I, *I, Robot*, Gnome Press, 1950.
[2] Balch, T. and Arkin, R.C.. "Behavior-based Formation Controlfor Multi-robot Teams." *IEEE Transactions on Robotics and Automation*, **14(6)**, 1998.
[3] Bruemmer, D. J., Dudenhoeffer, D. D., and McKay, M. D. "A Robotic Swarm for Spill Finding and Perimeter Formation," *Spectrum: International Conference on Nuclear and Hazardous Waste Management*, 2002.
[4] Chen, Q. and Luh, J.Y.S.. "Coordination and Control of a Group of Small Mobile Robots." In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2315-2320, San Diego CA, 1994.
[5] Crandall, J.W. and Goodrich, M. A., "Characterizing Efficiency of Human-Robot Interaction: A Case Study of Shared-Control Teleoperation." *Proceedings of the 2002 IEEE/RSJ International Conference Intelligent Robotics and Systems*, 2002.
[6] Kawamura, K., Peters, R. A., Johnson, C., Nilas, P., and Thongchai, S. "Supervisory Control of Mobile Robots using Sensory Egosphere" *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 2001.
[7] Fong, T, Conti, F., Grange, S., and Baur, C. "Novel Interfaces for Remote Driving: Gesture, Haptic and PDA," *SPIE Telemanipulator and Telepresence Technolgies VII* , 2000.

[8] Fong, T., Thorpe, C., and Baur, C., "Collaboration Dialogue, and Human-Robot Interaction," *Proceedings of the 10th International Symposium of Robotics Research*, 2001.

[9] Fong, T., and Thorpe, C. "Robot as Partner: Vehicle Teleoperation with Collaborative Control," *Workshop on Multi-Robot Systems, Naval Research Laboratory*, Washington, D.C, 2002.

[10] Fong, T., Grange, S., Thorp, C., and Baur, C., "Multi-Robot Remote Driving with Collaborative Control" *IEEE International Workshop on Robot-Human Interactive Communication*, 2001.

[11] Jones, C., and Mataric, M.J., "Sequential Task Execution in a Minimalist Distributed Robotic System" *Proceedings of the Simulation of Adaptive Behavior*, 2002.

[12] Parker, C. and Zhang, H., "Robot Collective Construction by Blind Bulldozing," *IEEE Conference on Systems, Cybernetics and Man*, 2002.

[13] Parker, L. E., "Evaluating Success in Autonomous Multi-Robot Teams: Experiences from ALLIANCE Architecture," *Implementations Journal of Theoretical and Experimental Artificial Intelligence*, 2001.

[14] Sheridan, T.B., *Telerobotics, Automation and Human Supervisory Control* MIT Press, 1992

[15] Simmons, R., Apefelbaum, D., Fox, D., Goldman, R. P., Haigh, K. Z., Musliner, D. J., Pelican, M., and Thrun, S., "Coordinated Deployment of Multiple, Heterogeneous Robots," *Proceedings of the Conference on Intelligent Robots and Systems*, 2000.

[16] Wang, P.K.C., "Navigation Strategies for Multiple Autonomous Robots Moving in Formation." *Journal of Robotic Systems*, **8(2)**, pp. 177-195, 1991.