

A Methodology to Assess Performance of Human-Robotic Systems in Achievement of Collective Tasks

Ayanna M. Howard

Jet Propulsion Laboratory, California Institute of Technology

4800 Oak Grove Drive, Pasadena, CA 91109, USA

howard@helios.jpl.nasa.gov

Abstract – In this paper, we present a methodology to assess system performance of human-robotic systems in achievement of collective tasks such as habitat construction, geological sampling, and space exploration. The methodology uses a systematic approach that assesses performance by incorporating capabilities of both human and robotic agents based on accomplishment of functional operations and effect of cognitive stress due to continuous operation by the human agent. In this paper, we provide an overview of the assessment system and discuss its implementation on a representative habitat construction task.

Keywords – Performance Assessment, Human-Robot Interaction, Task Allocation

1. INTRODUCTION

Typically, research that focuses on performance assessment of systems having both human and robotic agents tends to disregard the capability of one of the agents. In [1], a human-centered approach is used to understand the role of human-robotic teamwork in future human space exploration missions. In this work, a framework is developed in which robots become functional tools that assist the human rather than replace the human operator. In [2], the focus is to optimize overall performance by designing systems that use adjustable autonomy to dynamically change the autonomy of an intelligent agent. Different criteria are used to determine how the autonomy level, and thus the performance of the system, is adjusted based on reasoning about the costs of decisions. Recent work [3] has focused on evaluating human and robot teams through an analytical framework that decomposes tasks into independent functional primitives. Currently, the performance analysis proposed is in a generalized form that presents a concept of how to perform performance evaluation, but does not provide validated experimental results nor does it discuss the type of metrics needed for evaluation. In [4], complementary research is presented that introduces taxonomies and metrics useful for human-robot performance evaluation. In [5], Fong attempts to address the wide dispersion found in this area and develop common metrics for task-oriented human-robot interaction in terms of five task categories dependent on the level of human interaction.

Although research in human-robot performance assessment is expanding, approaches that integrate the contributions of both human and robot agents have been minimally addressed. We attempt to address these limitations by developing a systematic approach to assess system performance of human-robotic systems in achievement of collective tasks. The overall objective is to use performance characteristics to determine an optimal allocation of tasks to be divided between human and robotic system to minimize human mental workload while maximizing system performance, as necessary for such activities as habitat construction, geological sampling, and space exploration.

2. PERFORMANCE ASSESSMENT OF HUMAN-ROBOTIC SYSTEMS

In this section, we present a noninvasive method for performance assessment of human-robotic systems that evaluates the various effects of workloads on human performance and determines the performance tradeoffs derived from task allocation between humans and robotic systems. The approach is motivated by [6] and, as such, consists of four primary steps: 1) decompose scenario into a set of major functional task primitives and define performance metrics for each primitive, 2) evaluate the performance of all agents (human, robot) in performing each task primitive, 3) calculate a performance score based on satisfaction of task primitives and effect on agents, and 4) compute a composite task score to evaluate system performance during task achievement.

a. Performance Metrics

Workload studies are used to characterize human performance in terms of total demand placed on a person implementing a task. Developing a methodology to assess workload, or cognitive stress, using actual human subjects is a time consuming process, which must adequately deal with the inherent discrepancies found in the different subjects. To address this limitation, research efforts have focused on developing workload assessment models without the use of human subjects [7]. These efforts focus on decomposing tasks into a series of subtasks and assigning workload values by pairwise comparing the level of effort required to implement each subtask. Following this approach, we first decompose human-robot scenarios into a set of functional task primitives,

i.e. activities that need to be implemented by the human or the robotic system for goal achievement. In other work [3, 6, 8], an inclusive set of functional primitives in various space-related scenarios was constructed for assessing system performance. Using this as a basis, we constructed an elementary set of functional primitives, and identified the cognitive skills associated with each. To identify cognitive skills, we used the cognitive architecture construct [9] that breaks human information processing into 3 macro-level mechanisms: perception, cognition, and motor activities. Primitives were then selected to be as independent from each other as possible and to emphasize different aspects of perception, cognitive, and motor skills associated with mental demand (Table I).

TABLE I: Elementary functional primitives and associated activity type

Primitives	Primary Activity Type		
	Perception	Cognition	Motor
Grasp/Release			X
Identify	X		
Lift/Unload			X
Locate/Localize	X		
Mate/Unmate			X
Model/Represent		X	
Plan		X	
Track	X		
Traverse			X

By utilizing this elementary set of functional primitives, various scenarios can be defined by linking these primitives into a primitive hierarchy (or tree), which provides a broad understanding of the cognitive skills/mental demands required for each scenario. Each scenario is decomposed into its lowest level, such that the last node (or leaves) of the primitive hierarchy consists solely of the functional primitives identified above. As an example, Figure 1 depicts a branch of a primitive hierarchy for a habitat construction task (useful for Mars human-robot missions).

TABLE II. Associating performance metrics with functional task primitives for visual identification

Workload Value (0-10)	Performance Score (0-10)		Task Primitive
	Human	Robot	
0.0	10.0	10.0	No Visual Activity
1.0	10.0	8.5	Identify/Locate specific target
5.0	10.0	6.0	Localize with respect to target
5.4	7.0	5.0	Visually Track / Follow Target

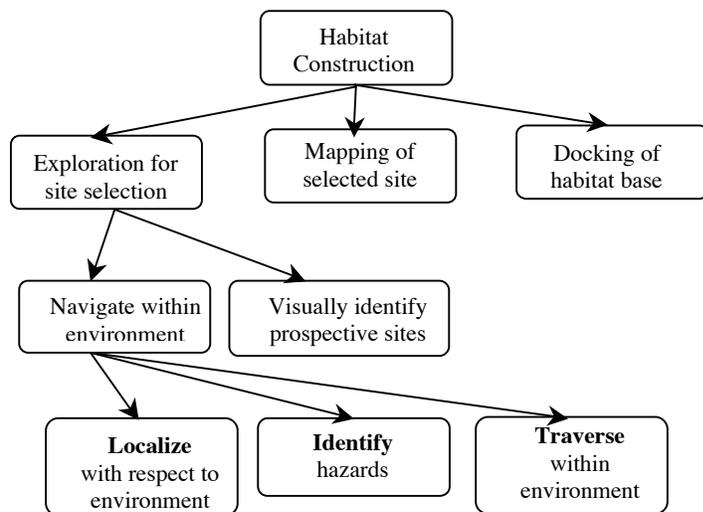


Figure 1. Primitive hierarchy for habitat construction task

Performance metrics are associated with the functional primitives located at the lowest node of the primitive hierarchy. Performance metrics consist of both workload value and performance scores, and use relative measures for human and robotic agents. Workload values are used to determine the relative decline in performance associated with the constant mental/resource load required to complete the task while performance scores quantify how well the agent achieves the task. Performance metrics are calculated for each elementary functional primitive and use relative measures in the [0,10] range for human and robotic agents based on a pairwise comparison method. To calculate human performance metrics, values are extracted by pairwise comparing the level of effort required to implement each subtask based on the calculated performance values. Performance scores for robotic systems are determined from evaluation of current robotic systems implemented in real-time [8]. As an example, the visual identification primitive in Figure 1 can be decomposed into the functional primitives and associated performance metrics as depicted in Table II. In this example, the performance score is ranked based on the minimum time necessary to complete a task, where 4.0 is baselined as the minimum performance value associated with a human agent. In our example application presented in Section 3, the performance scores are determined and compared to human and robot agents performing actions in the real world.

b. Performance Evaluation

We utilize the concept of an optimization function for dynamic task allocation [10] to calculate a composite task score using the detailed functional decomposition of a task scenario. The optimization function is dynamic in that it

incorporates both attributes of workload values and performance scores, which depends on the amount of time elapsed during real-world implementation. To determine the effect workload has on performance, we examined the work of Dinges and Mallis [11] in which studies were performed to determine the relationship between physical activity and performance of a human operator executing complex tasks over time. Using the resulting data from these studies, we mapped a logarithmic function to obtain a time-dependant performance trend associated with human task implementation. This trend reflects the effect of workload in our optimization function. For each task scenario, a composite task score can thus be constructed to determine overall system performance while incorporating the decreases in performance associated with consistent work operation, such that:

$$\forall \text{primitive} \in [1, n]$$

$$\rho = \text{PerformanceScore}(\text{agent}) \quad \text{agent} = \{\text{human}, \text{robot}\}$$

$$\omega = \begin{cases} s * \ln(s * \text{workload}) / \text{workload}, & \text{agent} = \text{human} \\ 0, & \text{agent} = \text{robot} \end{cases}$$

$$\text{Composite Task Score (s)} = \sum_{i=1}^n (\rho_i - \omega_i)$$

where s designates the repetitive number of scenario runs that have occurred, workload is the workload value associated with primitive i , and PerformanceScore is the performance score associated with primitive i . The composite task score is summed over all functional primitives for the task scenario and can be calculated for each repeated scenario run. A final composite task score provides an overall evaluation of relative performance for the scenario.

3. IMPLEMENTATION AND ANALYSIS

It is envisioned that future planetary exploration missions will involve humans and robots working in collaboration to accomplish both scientific and exploration goals [12]. To enable a long-term human presence in space, supporting technology, needed for tasks such as the capability for habitat construction, in-space assembly, and geological sampling, must be developed to enable the goals of these missions. One of the first steps in this process is to determine how to synergistically use humans and robots together in a systematic fashion. To validate the performance assessment methodology, we use our approach to assess the performance of human and robotic systems performing a simulated habitat construction task.

Our test environment (Figure 2) consists of a graphical user control panel that enables the human operator to control a robot operating in the real world. For this environment, we utilize the Sony ERS-210 robot for task implementation. The

control panel allows the human operator to view the world through the robot's eyes, as well as command the robot to move forward, backward, and turn either left or right. The human operator can also toggle between tele-operated control or autonomous behavior of the robot.

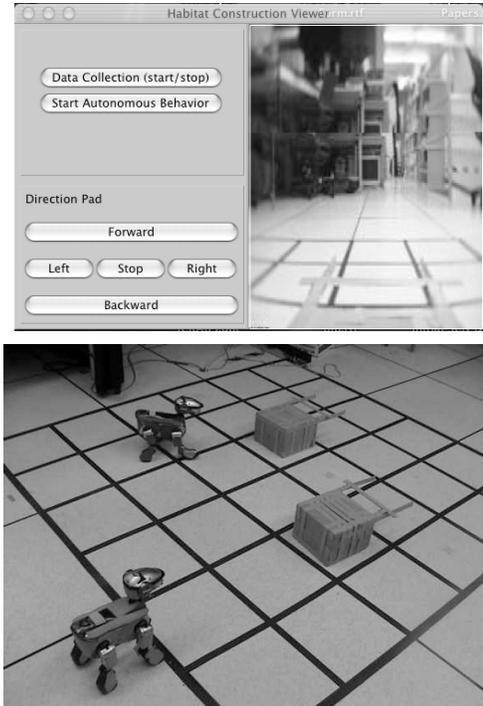


Figure 2. Test environment consisting of human operator unit and two Sony Aibo robots

The first step in the performance assessment process is to decompose the habitat construction task into functional primitives, and associate relative performance scores and workload values to each primitive. As our focus is on documenting the applicability of the assessment process to a representative task scenario, we use a simplified decomposition consisting of one branch, and two primary operations: locating the platform base unit in an obstacle-free environment and navigating to a position for subsequent transportation of the base units into a desired configuration (Figure 3). Figure 4 displays the output of the functional decomposition process, and provides the associated metrics used in the evaluation. For our current analysis, we wish to directly compare the performance of human tele-operator versus robot agents in the habitat construction scenario. The two set-ups we constructed for assessment were A) direct tele-operated control of the robot by a human operator (via the graphical user interface) and B) fully autonomous control of the robot, without direct human intervention. The autonomous behavior [13] programmed onto the robot allows the robot to search for and locate the platform base unit within the environment, and navigate toward the corresponding goal position for subsequent transportation of the base unit.

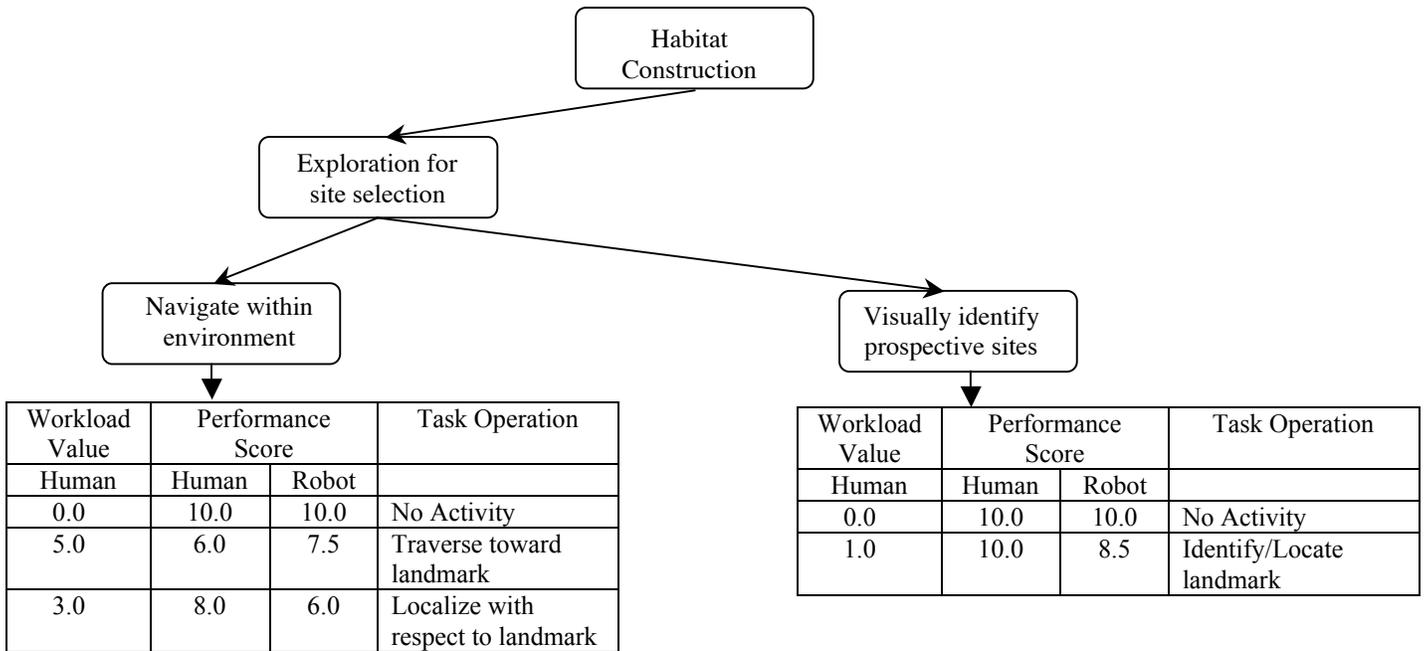


Figure 4. Decomposing the habitat construction task for extraction of metrics

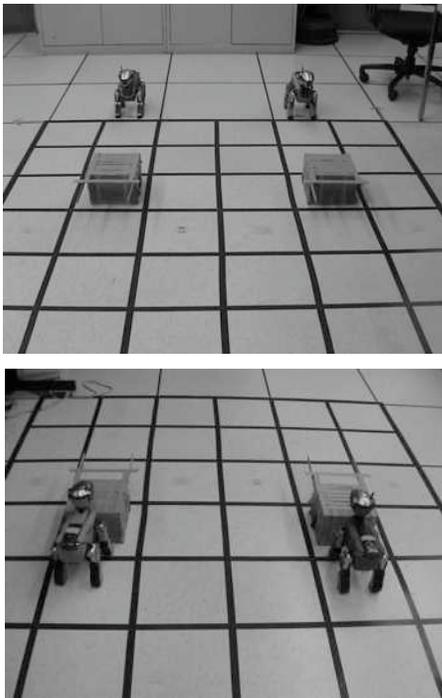


Figure 3. Initial and final robot configuration for habitat construction scenario

This involves implementing a vision-based algorithm to locate the base unit via color information, and extracting size and pixel location information from the image data. This information is then fed into a stored table that associates the two extracted image parameters with 3-D world position to which the robot is directed.

Figure 5 documents the composite task score calculated for each set-up based on the performance metrics and workload values shown in Figure 4.

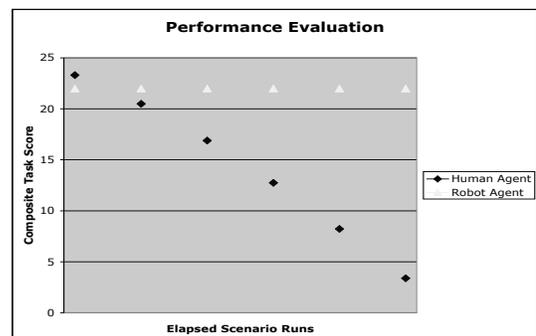


Figure 5. Calculated composite task scores

To compare our evaluation results with real world implementation, we ran through each scenario 10 times (with the habitat base unit located at different sites) for 6-10

continuous scenario runs and documented the execution time. To map execution time to composite task score, we correlated the elapsed time steps and scaled the execution times to match with the composite task score calculated for the robot agent. The time steps were selected to begin after the learning cycle for each scenario run (typically the first 2-4 runs). This process is acceptable because we are interested in understanding the *relative* performance of humans versus robots, and in capturing the corresponding decline in human performance associated with workload during real-time implementation. Implementing this normalization process gave us the outcome depicted in Figure 6 from a sample run of median error.

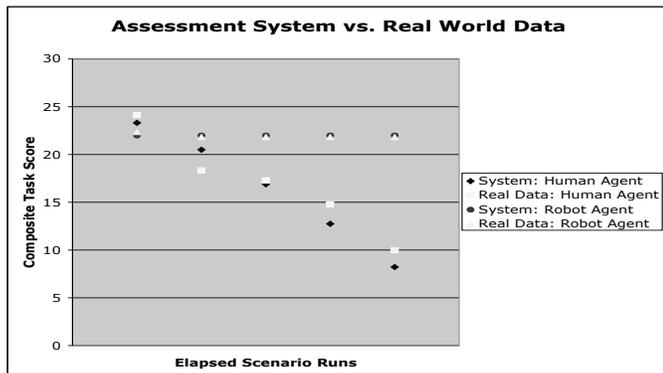


Figure 6. Comparison of performance assessment system versus real-world implementation data

As shown in Figure 6, the relative trend displayed by the performance assessment system compares favorably to the actual performance data collected during real-time implementation. As time elapses, the time for task completion by the human agent increases in the real-world implementation, while the task score calculated by the performance assessment system decreases. What we are interested in noting is that the performance assessment system is able to reflect the decline in performance during the real-time implementation process, and show the relative benefits of tele-operated and fully autonomous control. In essence, systems that evaluate human-robot interaction systems must incorporate aspects of both human and robotic system performance, i.e. the capability of the robotic system to implement tasks should be understood, as well as the human's ability to perform.

4. LIMITATIONS

The current version of the performance assessment system uses the pairwise comparison method to determine performance scores and workload values. This assumes ideal operating conditions and limits the ability of the system to handle unplanned discrepancies, such as extreme environmental complexity in the task space or untrained human operators. Future work for the assessment system will thus involve learning from the actual implementation data and

allowing refinement of the performance scores and workload values in real time. In addition, the performance scores and workload values incorporate crisp value definitions, and do not use relative ranges or allow for overlapping ranges in performance or workload. To allow full evaluation, this limitation can be resolved by incorporating a linguistic, or crisp intervals, for determining the composite task score. Lastly, the system assumes that the human agent is an expert in implementation of the task operations and does not incorporate the learning cycle required for a human operator to first become efficient in a new task. Future work will thus involve incorporating a parameter to acknowledge the learning lag necessary to correlate with real-world performance.

5. CONCLUSIONS

This paper presents a methodology to assess system performance of human-robotic systems in achievement of collective tasks. The methodology uses a two-tier process involving performance metrics and performance evaluation, which can be applied to a wide range of human-robotic activities performed in complex environments. The overall objective of the system is to use performance characteristics to determine an optimal allocation of tasks to be divided between human and robotic system to minimize mental workload while maximizing system performance. We have discussed the performance assessment methodology in detail and compared its implementation on a representative habitat construction task. The implementation of the method is shown to provide a correlated comparison that reflects the actual performance of human-robotic systems operating in the real world.

6. ACKNOWLEDGEMENT

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration and funded through the Draper Laboratory Research and Development Program. Thanks are due to Aryen Moore-Alston for managing the data collection process.

REFERENCES

1. M. Sierhuis, et. al, "Human-Agent Teamwork and Adjustable Autonomy in Practice," Proc. 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS), Nara, Japan, 2003.
2. P. Scerri, D. Pynadath, and M. Tambe, "Towards adjustable autonomy for the real-world," Journal of AI Research (JAIR), 2002, Volume 17, Pages 171-228.
3. G. Rodriguez, C. R. Weisbin, "A New Method to Evaluate Human-Robot System Performance," Autonomous Robots, 14(2-3), pp. 165-178, 2003.

4. H. Yanco, J. Drury, "A taxonomy for human-robot interaction," Proc. of the AAAI Fall Symposium on Human-Robot Interaction, pp. 111-119, 2002.
5. T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, A. Steinfeld, "Common metrics for human-robot interaction," in submission, 2004
6. A. Howard, et al., "A methodology to determine impact of robotic technologies on space exploration missions," Proc. of the 10th International Symposium on Robotics and Applications, Seville, Spain, June 2004.
7. C.D. Wickens, "Resource Management and Time Sharing." In Human Performance Models for Computer-Aided Engineering, NRC National Academy Press, Washington, D.C. 1989.
8. A. Howard, G. Rodriguez, "Validating Mission Relevance of Autonomy Technologies through Increased Science Return", Proc. 20th International Conference on Machine Learning, Washington D.C., August 2003.
9. W.W. Zachary, J.M. Ryder, J.H. Hicinbothom, "Building Cognitive Task Analyses and Models of a Decision-Making Team in a Complex Real-Time Environment," Cognitive Task Analysis, New Jersey, 2000.
10. B.P. Gerkey, M.J. Mataric "A formal analysis and taxonomy of task allocation in multi-robot systems," The International Journal of Robotics Research, Vol. 23, No. 9, 939-954, 2004.
11. D.F. Dinges, M.M. Mallis, G. Maislin, J.W. Powell, "Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management," U.S. Department of Transportation, Report No. DOT HS 808 762, 1998.
12. "President's Commission Report on Moon, Mars and Beyond: A Journey to Inspire, Innovate and Discover," Washington D.C, U.S. Government Printing Office, June 2004.
13. A. Howard, H. Seraji, "Vision-Based Terrain Characterization and Traversability Assessment," Journal of Robotic Systems, 18(10), pgs. 577-587, 2001.