2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)
Virtual Conference, July 8-10, 2021

Agility Metrics in the ARIAC Competition*

Anthony Downs[1], Zeid Kootbally[2] and Craig Schlenoff[1]

2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO) | 978-1-6654-4953-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ARSO51874.2021.9542831

Abstract— **A thriving manufacturing sector is the essential heart of a vibrant and balanced economy in the United States (U.S.). Small and medium-sized manufacturers (SMMs) constitute an important sector in the U.S. manufacturing but they are currently facing increasing competition due to economic globalization. To survive and thrive in this highly competitive environment, SMMs have to rely on automation and robotics, which bring with them a whole series of techniques to improve the quality and productivity of a manufacturing process. However, robotic systems need to be agile for them to be useful to SMMs so they can offer more automated customization of high-mix/low-volume production. This paper focuses mainly on the metrics used in the Agile Robotics for Industrial Automation Competition (ARIAC). The goal of the competition along with its associated metrics is to promote advances in research by assessing the performance of industrial robotic systems in manufacturing settings.**

I. Introduction

Agility, in the context of this paper, refers to the ability for robots to think, learn, and adapt in order to respond to failures during production. Among small-and-medium sized manufacturers, improving agility for the robots to perform a variety of tasks and be re-tasked randomly would be beneficial in a manufacturing process. To overcome agility challenges, focus is needed on key areas such as failure identification and recovery, automated planning, fixtureless environment, and plug-and-play robots. The current state of robotics shows that robots need to become more agile to support quickly changing requirements in their environment. Knowledge-enabled robots are needed that can execute their tasks with minimal upfront programming. Instead of having to reprogram robots when something changes, robots should be able to identify and accommodate the change. Robots need to have situational awareness, which involves detecting, identifying, and tracking objects and humans in their surroundings. Until robots can identify objects in the environment, and characterize the objects (what these objects are good for and how much they weigh), these robots cannot really do much. Robots, based on their capabilities and surroundings and a knowledge of what they are trying to accomplish, need to be able to plan how to assemble parts and build kits. This is the type of knowledge that we would like robots to have, so they can understand what they can do.

To advance the agility performance of manufacturing robotics assembly systems in unstructured and dynamic environments, the Agile Robotics for Industrial Automation Competition (ARIAC) was initiated in June 2017 by the National Institute of Standards and Technology (NIST) in collaboration with the Open Source Robotics Foundation (OSRF). ARIAC is a simulation-based competition to allow competitors around the world to utilize latest advances in artificial intelligence and robot planning to address real-world industrial challenges pertaining to kitting, assembly, and order fulfillment applications. The latest iteration of the competition was held between April and May 2020 and introduced a more challenging environment, a new robot, new scenarios, and new agility challenges. The annual occurrence of ARIAC is two-fold. First, NIST intends to use the results and knowledge gained from ARIAC to further its efforts to develop metrics and test methods to measure robot agility as well as tools for manufacturers to assess the agility of their robotic systems. Second, ARIAC aims to encourage competitors to develop the most effective solutions to real world industrial robot challenges, enabling the greater use and productivity of robotic systems by industry.

Within ARIAC, the competitors are required to develop a robot control system to perform either kitting, assembly, or order fulfillment in a simulated environment. Gazebo[1], which is an open source robotics simulation environment, is used as the testing platform. The Robot Operating System (ROS)[2], which is an open source set of software libraries and tools, is used to define the interfaces to the simulation system. Gazebo was chosen because it is commonly used in academia. Additionally, it is free and, therefore, no mandatory monetary investment is required to compete.

The competition addresses the aspect of robot agility that focuses on software, including knowledge representation, planning, and decision making. While hardware aspects (such as different types of grippers) can play a large role in agility, they are not the focus of this competition. Perception and grasping have played a minimal role in the competition to date, but are expected to increase in importance in future years.

Competitors were faced with challenges such as forced dropped parts and in-process order changes [3]. Each competitor's system had to address these challenges and attempt to finish the goal autonomously in real-time. The scoring metrics used in the competition were partially based on the

*Certain commercial products or company names are identified here to describe our competition. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products or names identified are necessarily the best available for the purpose.

[1]Anthony Downs and Craig Schlenoff are with the Engineering Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, MS 8230, Gaithersburg, MD 20899, USA `anthony.downs@nist.gov`, `craig.schlenoff@nist.gov`

[2] Zeid Kootbally is with the Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA 90089, USA `kootball@usc.edu`

robot agility metrics developed by NIST [4]. Competition scoring took into account whether the goal was completed (both quantitative and qualitative metrics), how fast the goal was completed, and the cost of their sensor configuration. Each competitor's system was given a cost based on the number and type of sensors used. Typically, sensors that gave more information were priced higher than sensors that gave less. The total cost of each competitor's configuration factored into their score; cheaper configurations resulted in higher scores for the that element of the overall score.

The rest of the paper is organized as follows. Section II describes different robotic competitions along with some of the metrics involved in those competitions. Section III details the metrics that were used in ARIAC and the evolution of those metrics over the years of the competition. Section IV describes how those metrics were implemented in the ARIAC environment. Section V describes the future of ARIAC and how we foresee the metrics and the competition evolving.

II. Performance Metrics in Robotic Competitions

With the emergence of advanced technologies, agile manufacturing techniques have gained considerable popularity to boost productivity, effectiveness, responsiveness, and product quality. Agile manufacturing is regarded as a new concept to respond to the dynamic and fast-changing environment [5] and is described as a crucial characteristic for manufacturing companies to maintain their competitiveness [6]. Therefore, there is a need for performance measures to characterize and compare robots and robotic systems to help determine which features are most suited to a particular application [7]. Robotic systems are complex and involve a wide range of features and performance characteristics whose importance differ depending on the application domain.

Performance metrics are usually difficult to define because the requirements on which the domain is based can be changed according to the user's needs. One of the particulars of the metrics is that they should be practical and constructed to expand the community interest. Performance metrics should also be independent of the software and hardware. For instance, metrics used in simulation to assess the performance of a gantry robot should be applicable to a real robot mounted on a linear rail, as long as the scenarios are otherwise the same.

Many robotics competitions have been held over the past decade. These competitions have often had the goal of comparing different robotic systems and their research approaches. Comparing robotic systems is one way to ensure that the correct robotic system is used for the correct application. When designing the rules for a competition, there are several ways to compare the performance of robotic systems [8].

While many competitions involving students fall in the categories of subjectively-ranked and non-ranked competitions, competitions that promote advances in research and seek the most efficient systems are classified in the objectively scored competitions category. Some of these objectively scored competitions and their metrics are described below.

The Amazon Picking Challenge [9] was a yearly competition from 2015-2017, focusing on "picking". In the competition, teams had to develop robotics hardware and software that can recognize objects, grasp them, and move them from place to place. The goal was to use this competition to assess if robots would be able to do some of the menial pick and place operations that are currently performed by humans. The scoring rubric was mainly related to picking a target from bins with bonus points awarded for items that were difficult to grasp. Points were lost for damaging any item, picking the wrong item (and not putting it back), or dropping the target item anywhere but into the destination tote.

In the Virtual Defense Advanced Research Projects Agency (DARPA) Robotics Challenge, which ran from June 17-21, 2013, teams competed in a simulated suburban obstacle course. Twenty-six teams from eight countries qualified. Competing teams applied software of their own design to a simulated robot in an attempt to complete a series of tasks that were prerequisites for the next stages of the grand challenge [10]. The overall DARPA Robotics Challenge, which included both the virtual and physical challenges, was launched in response to a humanitarian need that became glaringly clear during the nuclear disaster at Fukushima, Japan, in 2011. The challenge organizers leveraged the work done by NIST on ASTM Committee E54.08.01 rescue robotics metrics [11]. Instead of scoring the functions that describe robot performance (e.g., visual acuity, dexterity, maneuverability), the organizers looked at holistic solutions that reflect a robot's ability to complete a mission.

Robot Competitions Kick Innovation In Cognitive Systems and Robotics (RoCKIn) is a European Union funded project aiming to foster scientific progress and innovation in cognitive systems and robotics through the design and implementation of competitions. RoCKIn@Work [12] [13], a subset of this competition, looked for innovative industrial robots that could help businesses meet increasing demand from their customers. The competition proposed different benchmarks to assess a robot in three categories: Object perception, object manipulation, and control.

III. Performance Metrics in ARIAC

When the initial version of ARIAC was being developed between NIST and OSRF, the NIST organizers searched through the literature to find applicable metrics for measuring the agility of industrial robots. These metrics needed to be able to both compare different systems as well as different configurations of systems where alternative choices were made to improve the robot system's agility. To better understand the descriptions provided in the following sections, the reader may refer to the following terms in the context of the competition: A **trial** in this context is a single run of the simulation in which at least one order is described. An **order** is a goal state and consists of information on parts to build kits. The order also specifies where to build kits, which is usually on one of the two automated guided vehicles (AGVs) present in the environment. An order consists of at least one **shipment**, where a shipment is an instance of an order. A **kit** is the result of a process which groups separate but related items (parts in ARIAC) as one unit. Kits are built in kit trays located on AGVs. A **shipment** is an instance of an order. If an order must be built and delivered multiple times, then the order consists of multiple shipments.

Current Metrics for ARIAC

The most recent form of the competition (ARIAC 2020) [3] has three generalized metrics that were developed to measure the agility performance of the robotic systems. These three metrics are Cost Factor, Completion Score, and Efficiency Factor. Each of these individual metrics can be used for a standalone comparison of the systems on that individual scale, but they are also combined for the purposes of ARIAC with a set of constant factors to provide a single score for ranking the competitors overall.

A. Cost Factor

The Cost Factor is based roughly on the cost of the overall system components. As a general metric this would include everything from the robot to the mountings to the sensors that allow the system to observe the environment around it.

Within ARIAC, the competitors are limited in their choice of robot or robots depending on the theme of the year, and the mounting choices are also set by the competition organizers. So the Cost Factor in these cases gets limited to the choice of sensors placed and used in the environment by each competitor. Each of these sensors is given a nominal cost by the organizers in order to encourage different modes of agility. The Total Cost (TC) representing the sensors chosen by the competitors is represented by Equation 1 where the total number of sensors is n.

$$TC = \Sigma_{i=1}^{n} Cost_i \tag{1}$$

The organizers set a Baseline Cost (BC) as a representative set of sensors for comparison purposes, such that the expected costs from the competitors will have some above the baseline and some below. The Cost Factor (CF) is then calculated using Equation 2.

$$CF = \left(\frac{BC}{TC}\right) \tag{2}$$

B. Completion Score

The Completion Score is a metric measuring how well the competitors are able to complete the kits required for the particular order, with the goal of rewarding a fully assembled kit being submitted over a partial or wrong kit. For a given order for submission, S_j, which contains i parts, the following points would be available:

- 1 point (up to i points) for each part of the correct type placed in the kit tray.
- 1 point (up to i points) for each part placed in the correct position ($\pm 3\,cm$) and orientation ($\pm 0.1\,rad$).
- 1 point (up to i points) for each part of the correct color placed in the kit tray
- an additional bonus of i points if all three categories above received the maximum score.

C. Efficiency Factor

The Efficiency Factor, EF, measures the efficiency of the system by comparing the time to complete a task for a system (from one competitor) with the average of all systems (all competitors) performing the task. When an order is sent to the competitors, a timer is started that will end when the order is declared complete and delivered or when a time limit is reached. For a trial j, each competitor has their time T_j, which gets averaged together as $AT_j = avg(T_j)$. Note that if a competitor's system times out (taking longer than 500 simulation seconds), the efficiency factor is set to 0 and the trial's time is not included in the average. The Efficiency Factor is then calculated as shown in Equation 3.

$$EF_j = \left(\frac{AT_j}{T_j}\right) \tag{3}$$

Trials with a changeover, i.e., when there is a higher priority order sent to the competitors during execution, have separate timers for each order that is sent to the competitors.

D. Constant Factors and the Trial Score

The three metrics described above are combined with constants, which are set by the organizers, to calculate the Trial Score, TS. The cost factor is combined with the average of the completion scores across however many kit orders are in the trial. The efficiency factor for each kit order is combined with the completion score for that order, and in the trials where there was a higher priority order sent, a high priority constant factor ($h = 3$) applies a higher bonus for completing that high priority order faster. The Trial Score for a single trial is calculated using Equation 4.

$$\begin{aligned} TS = {} & (CF \times AVG(CS)) \\ & + (EF_1) \times (CS_{S_1}) \\ & + h \times (EF_2) \times (CS_{S_2}) \end{aligned} \tag{4}$$

The Trial Score for each trial is added for each competitor and then points are awarded to competitors based on their rank. The team with the highest total TS is awarded 80 points, while the second highest gets 70, followed by 60, and so on. If there are more than 8 teams that complete the finals, the remaining teams are awarded 0 points and get a score solely based on the judging panel as described below.

E. Judging Panel

Starting in the 2nd year of ARIAC (ARIAC 2018), in order to provide a human judgement component to the competition scoring, as well as to satisfy a requirement of the prize contest portion, a panel of three human judges was chosen from representatives from industry in order to be the subjective judgement making up the last 20 points possible for the overall score. Each judge is tasked with evaluating the competitors' performance individually in terms of both innovativeness and feasibility of the approach. The judges are able to watch video playback of the competitors' performance via a series of highlight videos provided by the competition organizers. For innovativeness, the panel starts with a default score of 0 and awards points for how the competitors show that they have an innovative approach to the scenarios, up to a maximum of 10 points. For feasibility, the panel starts out at 10 points possible and deducts points based on the individual judge's subjective evaluation of how feasible the solutions would be to implement in an actual manufacturing plant. The judging panel's innovativeness and feasibility scores are added together and then averaged across the judges to get the final possible points for the overall score.

IV. Implementation

As described in the previous section, the current version of ARIAC consists of three main metrics, namely Cost Factor, Completion Score, and Efficiency Factor. The Cost Factor and the Efficiency Factor are not computed during competition runs (or trials) but rather after the trials end. This section describes the approach used to compute each one of these metrics either during or after trials.

A. Trial Infrastructure

All ARIAC events are run through GEAR (Gazebo Environment for Agile Robotics), a software originally developed by OSRF and now maintained by NIST. The GEAR interface allows for a controlled standardized means of communication between competitors and the simulation environment. To maximize flexibility, GEAR was implemented to be a ROS-based interface. While new features were added to GEAR for each iteration of ARIAC, its structure has remained consistent across competitions. With GEAR, competitors implement their system in a variety of supported programming languages. Additionally, this approach was chosen to isolate the use of a simulated environment as an implementation detail. Competitors' systems never communicated directly with the Gazebo simulator, but instead, with GEAR which in turn communicated with the simulator via a Gazebo-ROS integration layer. Correctly-designed kitting systems developed to work in a simulated environment should be usable on a physical robot with minimal software modifications due to the use of an abstract ROS interface. Similarly, kitting systems developed to control a particular manipulator can be used to control another manipulator with minimal modifications if designed appropriately.

Through the GEAR interface, the NIST organizers can control the type, the color, the quantity, and the location of parts to be used in a trial. Other components, such as the kits to build and the agility challenges to use during a trial, are controlled by GEAR.

A typical trial consists of the chain of events depicted in Figure 1. Through a ROS Service, competitors first need to start the competition, which in turn activates different components of the GEAR interface, such as starting the conveyor belt or moving the robot to its home position. Starting a competition also allows GEAR to evaluate each shipment submitted during a trial. Once the competition has started, competitors will start receiving orders through a ROS Topic. Kitting is then performed to reach the goal state as described by the order. During kitting, low-level or/and high-level challenges may be triggered. Competitors' systems must handle these agility challenges to receive the maximum points allocated for the current trial. In this context, low-level and high-level agility challenges are defined as follows:

- Low-level agility challenges require immediate interactions between the robot and the parts. For instance, one of the low-level agility challenges consists of the robot placing faulty parts in a kit tray. A camera above the kit try notifies the robot about faulty parts. The robot has to discard these faulty parts.
- High-level agility challenges consist of challenges which require planning and scheduling. For instance, some challenges require the robot to do path planning and use forward kinematics to avoid moving obstacles to reach a certain location in the workcell.

When shipments are ready, competitors' systems need to submit these shipments for evaluation. When all the orders have been fulfilled, the competition must be ended by competitors' systems through another ROS Service call.

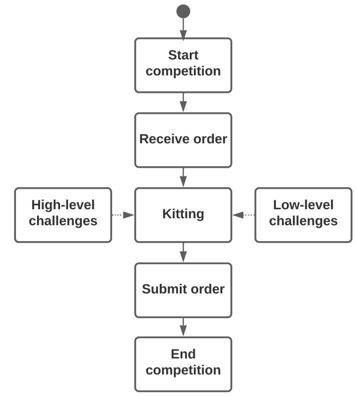

Fig. 1: Flow diagram showing the chain of events that are performed by a competitor's system during a trial. Note that there is a possibility that low-level and high-level challenges may not occur at all during a trial.

B. In-trial Evaluation

The Completion Score is computed via ROS plugins within the Gazebo simulation environment after each shipment is submitted, as shown in Figure 2. Note that the Completion Score for each shipment is summed up to get the Completion Score for the order. First, GEAR performs an analysis of the submitted shipment and validates this shipment against the expected shipment. There are two situations in ARIAC which automatically nullify the Completion Score for the shipment: 1) If there is any robot-robot collision (i.e., robot arms colliding with each other) or any robot-human collision and 2) if the shipment is submitted using the wrong AGV. If none of these two situations is encountered, GEAR proceeds with computing the Completion Score for the shipment. For each product (part) in the shipment, GEAR awards 1 point for correct product type, 1 point for correct product color, and 1 point for correct product pose. Only if

TABLE I: Cost Factor computed in the finals of ARIAC 2020.

Teams	Sensor Types and Numbers						Total Cost (TC)	Cost Factor (CF)
	Logical	RGBD	Depth	Break-beam	Laser Profiler	Proximity Sensor		
Team1	16	0	0	4	0	0	8400	1.19047619
Team2	17	0	0	8	0	0	9300	1.075268817
Team3	15	0	0	8	0	0	8300	1.204819277
Team4	9	0	0	8	0	0	4600	2.173913043

all these three situations are true then the all product bonus (1 point) is awarded for the product. Once each product within the shipment has been processed, the Completion Score is computed for the shipment. If more shipments are expected for the current order then GEAR waits until either the other shipments are submitted or the trial has reached the time limit (500 simulation seconds). At the end of the trial a score breakdown is provided for the current order.

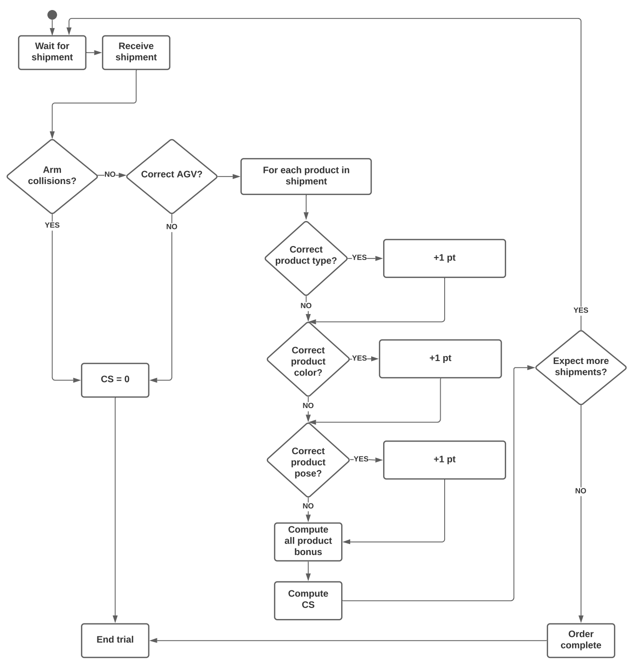

Fig. 2: Flow diagram showing the different steps used to compute the Completion Score for one shipment.

C. Post-trial Evaluation

While the Completion Score is calculated after each shipment during a trial, the Cost Factor is computed offline, i.e., outside trial runs. To compute the Cost Factor, the organizers first gather the type and the number of each sensor used for each competitor and store this information in a spreadsheet. As competitors have to use the same set of sensors for all the trials, gathering sensor information is performed only once for each competitor. The Cost Factor is then computed using Equation 2. Table I shows the Cost Factor for the finalists in ARIAC 2020. For privacy reasons, real competitors' names have been replaced with a team number (e.g., Team1). While most competitors mainly used logical sensors, not all the competitors used the same number of logical sensors. This is due to competitors' ability and flexibility to mount sensors anywhere in the environment and each competitor has their personal choice and strategy. From Table I, for example, Team4 used a very small number of logical cameras compared to the other competitors. This is translated into Team4 having the largest Cost Factor among the finalists. In contrast, Team2 has the lowest Cost Factor since the Total Cost for Team2 is the largest among the finalists.

The Efficiency Factor (see Equation 3) and the Trial Score (see Equation 4) are both computed offline. After a trial has completed, the organizers inspect ROS log files to gather information on the competitor's system for a given trial. The chart presented in Figure 3 displays the Trial Score for each competitor for each one of the 15 trials used in the finals. For each competitor, their Trial Scores are added, which resulted in a final score of 201.10 for Team1, 690.65 for Team2, 546.92 for Team3, and 410 for Team4. From these results, each competitor is then awarded a number of points based on the ranking, as described at the end of Section III-D.

V. The Future of ARIAC

To date, the competition and its metrics have mostly been developed and determined by the NIST organizers. However, there are plans underway, described in this section that will broaden the pool of the development of these metrics.

As the competition evolves over the years to adapt to different scenarios and themes, the details of the metrics get adapted to meet the needs of that year's competition. For example, the completion score metrics for the 2021 competition have been modified slightly to account for adding assembly into the mix. There will also be new metrics that are being developed through the Measuring Robot Agility Working Group. Starting in September 2020, the Measuring Robot Agility working group (under IEEE Standards Association (IEEE-SA), Robotics and Automation Society) [14] has been meeting every two weeks to develop a set of standard test methods and metrics for measuring robot agility. The working group is aiming to develop metrics for testing each of the following aspects of agility: hardware and software reconfigurability, communications, task representation, sensing & perception, reasoning, planning, tasking, and execution. As draft metrics become available from this working group, the ARIAC competition will be one of the venues for testing out these metrics to see how effectively they differentiate between robotic systems. Through this collaboration, the working group and ARIAC will each benefit from one another as the development continues. For more information

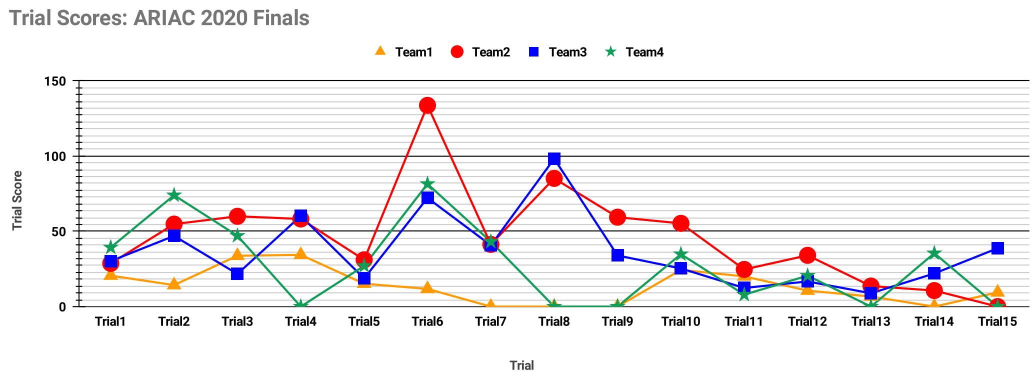

Fig. 3: Trial Score computed for each competitor's system in the finals for 15 trials.

on the working group or to get involved, please contact the authors.

The competition has gone through many changes through the years, transitioning from being co-developed by NIST and OSRF to being developed and maintained by NIST these past two years. It has also gone from a non-prize competition in 2017, to a prize competition beginning in 2018. The theme and scenarios of the tasks being given to the competitors have also changed through the years, from kitting and order fulfillment to a combination of kitting and assembly.

There are early plans in process to try aligning a future year's competition more closely with an industry partner. The plan is to have the winning solution(s) be more directly applicable to the partner in order to promote a closer tie between the research being done and the industry that will be using the solution(s).

VI. Conclusions

Throughout the five years of the Agile Robotics for Industrial Automation Competition (ARIAC), NIST has been able to use an online simulation competition structure to advance the agility performance of manufacturing robotics systems to promote the ability for robots to be able to sense, evaluate, and adapt to changing conditions in the environment. These changes will allow more (and smaller) manufacturers to start utilizing robots more often, freeing up their people to contribute directly to social efforts, and enable companies to differentiate on what social efforts they choose to contribute to. The metrics that were initially developed by NIST have served as the basis for measuring, ranking, and comparing the competitors performance in the competition. These metrics will be expanded upon in the near future with work from the IEEE-SA Measuring Robot Agility Working Group to further the development of standard metrics and test methods. The solutions and metrics that come out of the competition will shape the future of manufacturing robotics assembly systems.

ACKNOWLEDGMENT

Dr. Kootbally acknowledges support for this work under grant 70NANB19H009 from the National Institute of Standards and Technology to the University of Southern California.

References

[1] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3. IEEE, 2004, pp. 2149–2154.

[2] A. Koubâa, *Robot Operating System (ROS).* Springer, 2019, vol. 1.

[3] A. Downs, Z. Kootbally, W. Harrison, P. Pilliptchak, B. Antonishek, M. Aksu, C. Schlenoff, and S. K. Gupta, "Assessing industrial robot agility through international competitions," *Robotics and Computer-Integrated Manufacturing*, vol. 70, p. 102113, 2021.

[4] A. Downs, W. Harrison, and C. Schlenoff, "Test methods for robot agility in manufacturing," *Industrial Robot: An International Journal*, vol. 43, pp. 563–572, 08 2016.

[5] P. Ifandoudas and R. Chapman, "A practical approach to achieving agility – a theory of constraints perspective," *Production Planning & Control*, vol. 20, no. 8, pp. 691–702, 2009.

[6] H. Sharifi and Z. Zhang, "Agile Manufacturing in Practice: Application of a Methodology," *International Journal of Production Economics*, vol. 21, no. 5–6, pp. 772–794, 2001.

[7] M. Shneier, C. Schlenoff, F. Proctor, T. Kramer, J. Falco, and E. Messina, *Measuring and Representing the Performance of Manufacturing Assembly Robots*, 2015, no. NISTIR 8090.

[8] H. Yanco, "Designing metrics for comparing the performance of robotic systems in robot competitions," in *Proceedings of the Workshop on Measuring Performance and Intelligence of Intelligent Systems (PERMIS), IEEE Conference on Control Applications*, 2001. [Online]. Available: https://doi.org/10.1145/2393091.2393104

[9] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and Observations from the First Amazon Picking Challenge," 2016.

[10] DARPA, "DARPA Virtual Robotics Challenge Results," DARPA Virtual Robotics Challenge results, 2013.

[11] N. I. of Standards and T. (NIST), "Standard Test Methods for Response Robots," https://www.nist.gov/el/intelligent-systems-division-73500/standard-test-methods-response-robots, 2020.

[12] R. Bischoff and T. Friedrich and G.K. Kraetzschmar and S. Schneider and N. Hochgeschwender, "RoCKIn@Work: Industrial Robot Challenge," in *The Oxford Handbook of Innovation*. IntechOpen, 2017, ch. 4, pp. 47–73.

[13] E. Union, "RoCKIn@Work: Industrial Robot Challenge," http://rockinrobotchallenge.eu/work.php, 2017.

[14] I. E. E. E. Standards-Association, "Measuring robot agility working group (ieee-sa p2940)," https://standards.ieee.org/project/2940.html, 2020.