# Improving Robotics Competitions for Real-World Evaluation of AI

**John Anderson and Jacky Baltes**
Dept. of Computer Science, University of Manitoba
Winnipeg, Manitoba
R3T 2N2 Canada
*andersj,jacky@cs.umanitoba.ca*

**Kuo-Yang Tu**
Institute of System Information and Control
National Kaohsiung First Univ. of Science and Tech.
Kaohsiung, Taiwan
*tuky@ccms.nkfust.edu.tw*

### Abstract

While embodied robotic applications have been a strong influence on moving artificial intelligence toward focussing on broad, robust solutions that operate in the real world, evaluating such systems remains difficult. Competition-based evaluation, using common challenge problems, is one of the major methods for comparing AI systems employing robotic embodiment. Competitions unfortunately tend to influence the creation of specific solutions that exploit particular rules rather than the broad and robust techniques that are hoped for, however, and physical embodiment in the real world also creates difficulties in control and repeatability. In this paper we discuss the positive and negative influences of competitions as a means of evaluating AI systems, and present recent work designed to improve such evaluations. We describe how improved control and repeatability can be achieved with mixed reality applications for challenge problems, and how competitions themselves can encourage breadth and robustness, using our rules for the FIRA HuroCup as an example.

## Introduction

Situated and embodied approaches have been a strong influence on moving artificial intelligence from developing brittle techniques that operate only under significant assumptions in simplified environments, to developing robust solutions that operate in the real world. Chief among these situated, embodied approaches has been the influence of robotic applications. Robotic systems, like many other situated approaches, force AI researchers to focus on designing complete systems for a given environment, as opposed to isolated aspects that ignore the interaction of complex phenomena. A complete robotic solution requires consideration of all problems, from sensing and perception to planning and action, under the intended conditions of robot operation. Beyond simulated systems and other forms of embodiment however, robotic systems force a strong physical grounding, including elements of the real world that are impossible to simulate accurately, such as wheel slippage due to uneven surfaces and the many other sources of sensor error (e.g. everything from changing lighting conditions to dust in the air affecting cameras).

The added richness and complexity introduced by robotic embodiment greatly complicates the evaluation of intelligent systems. Part of this difficulty is the highly-interconnected nature of an intelligent system, its embodiment, and the peripheral systems it depends on. If a vision system is not working, for example, an intelligent control system will certainly not be performing at its best. A significant part of the difficulty in evaluation, however, is the embodiment itself: the same physical embodiment that forces researchers to consider all the complexities of the real world results in additional physical dependencies that make running experimental trials difficult: mechanical joints fail, sensors break, and batteries wear down during trials. The likelihood of an experimental trial being invalidated through some type of failure is especially high in large-scale multi-robot systems, which is why much of this type of research still relies heavily on simulators such as Player/Stage (Gerkey, Vaughan, & Howard 2003) for evaluation.

The failure of robots and their components, however, is just one facet of a more general problem in physical evaluation: control over the factors being examined under experimentation. It is difficult, for example, to properly place physical robots in precisely the same positions for multiple trials, or to properly randomize their positions in the physical world according to a statistical distribution and subsequently place robots accurately. Where events must occur in the real world (e.g. the likelihood of a collapsing floor in a USAR domain, or the likelihood of encountering a moving obstacle), it is equally difficult to create a consistent probability in a real-world domain. The physical measurement of moving robots can also be challenging. While all of these can be dealt with in simulation, reverting completely to simulation invariably removes elements that can only be properly accounted for in the real world. This results in an overly-optimistic view of the performance of any system, and also removes some of the very factors that drive research to improve such systems for the future.

In part as a reaction to the problems of repeatability and controlled experimentation in complex robotic domains, structured robotic competitions involving a common challenge problem through which to compare alternative approaches have become a common method of evaluating the performance of robotic techniques *in situ*. The challenge problem becomes a broad set of controls over the intended domain, and is supplemented with lengthy and precise rules intended to ensure a fair comparison between approaches. The most well-known of these challenge problems is soccer, used by the RoboCup and FIRA federations, both of which have operated using variations of this problem for over ten years. These competition-based models have led to a great deal of publicity and are widely used to attract students to AI and robotics, and to science, engineering, and technology in general.

While there continues to be discussion over the precise elements of competitions that best serve to motivate and educate (e.g. (Fox 2007)), and whether other approaches might be more suitable in some ways or to some specific audiences (Hamner *et al.* 2008), here we consider the use of competitions from an evaluation perspective. There are many positive elements that competitions bring from an evaluation perspective, and despite their popularity, challenges remain to be overcome. In this paper, we describe two pieces of recent work in improving competition-based evaluation of AI. The first of these, mixed reality, involves the addition of elements of simulation to physical robotics for evaluating intelligent control in classroom-based competition environments. This allows us far greater control over the environment for both competition and experimental evaluation than physical robotics alone, and also supports more interesting applications as well. The second of these involves the design of competition rules to promote breadth, flexibility, and adaptability. These principles are embodied in our design of the current FIRA HuroCup competition rules. We begin by discussing the main advantages and challenges associated with competition-based evaluation, to put this work into context.

## Robotics Competitions from an Evaluation Perspective

Competition-based challenges such as robotic soccer bring some very important elements of the real world to evaluating robotic systems, compared to laboratory evaluations. For example, there is generally much greater time pressure: a precise schedule must be adhered to, and robots must be able to perform their tasks at that time or default. While off-line development time may continue for many months prior to the competition given the pre-circulated rules, development time during the competition itself is strictly limited. Moreover, because the real world is involved, there will always be some situations that cannot be fully anticipated in the rules that must be adapted to on-site. Since the intent of the designers of the competition is to anticipate as much as possible in terms of defining the rules, these are usually small (e.g. the color of paint for a goal may be slightly different than described, or lighting intensity may differ), but it is always possible to have to adapt more significantly. Overall, this approach greatly encourages balancing the theoretical performance of the solutions developed for a challenge, with their brittleness, adaptability, and reliance on extensive resources (set-up time, manpower) in practice: all factors that must be adhered to in the real world.

Challenge environments also serve to prevent self-censorship bias in dissemination of scientific results. Since success stories are much more likely to be published than negative results, other researchers do not gain the benefit of learning from the failures of their colleagues. In a competition environment, approaches are compared head-to-head, and every competition thus results in dissemination of significant negative as well as positive results, greatly increasing the scientific rigour of the evaluation.

Challenge environments are also adaptable over time, to increase difficulty as the solutions of the participants improve based on solutions demonstrated in prior years. For example, various RoboCup leagues have made adaptations ranging from removing walls around a field, removing unique fiducial markers for localization, and relaxing lighting standards over the past ten years. In spite of the many possible constraints that affect how robot soccer is played, the game is still a reasonably restricted and well-structured environment compared to the bulk of human activity (e.g. localization is relatively simple on a playing field, just as it is in human soccer). Because of this, broader and less well-structured challenge problems have emerged, such as those embodied in NIST's urban search and rescue (USAR) test arenas (Jacoff *et al.* 2003). There, robots are expected to traverse unmapped terrain with significant challenges both physically (e.g. stepping fields, multiple levels) and perceptually (glass, mirrors, repeated visual patterns, moveable debris that will confound wheel-encoders), while looking for simulated human victims.

While these challenge domains are much closer to real-world evaluations than the vast majority of laboratory settings, they also introduce problems in robotic evaluation because of the precise rules they require. In humanoid soccer, for example, it is much easier to find and kick a ball with multiple cameras (e.g. at foot level), or omnidirectional vision, but these violate the principles of comparison to human abilities. Similarly, proportionally large feet provide more stability when walking while violating similar principles, and larger humanoid robots can cover ground faster than smaller ones. Rules must be in place limiting the degree to which these dimensions can be exploited, and there are an unending number of variants that must be considered and restricted to preserve a fair comparison. The effect of these highly precise rules on participants significantly effects the utility of the competition as a means of evaluation. Because everyone involved is focussed on the details of the particular challenge, there is a very strong motivation to improve performance by exploiting minutia in the rules governing the challenge, or thinking of aspects that have not yet entered the rules, rather than confronting the difficulties associated with the challenge directly. The result is instead of developing interesting, adaptive, flexible approaches, research groups are encouraged to create narrow, special-purpose solutions (and then to adapt that narrow solution to changes from year to year).

In robotic soccer for example, one of the reasons the game was chosen is that it allows for sophisticated teamwork, and real-time pressures for high-level control such as choosing plays and lower-level issues such as path planning and motion control. Considering the small-size league at RoboCup, however, the performance of any team relies much more on fast omni-directional drives and specialized devices such as dribble bars and chip-kickers. The latter of these devices effectively allows a robot to shoot the ball from any point on the field at a speed high enough for a goalie to have problems deflecting it or even moving toward the area of the scoring attempt, and goes far beyond what any human soccer player could be expected to do. This leads to teams developing highly specialized units for retrieving and launching orange golf balls, rather than the broad range of skills that typify a human soccer player. Moreover, these specialized devices will likely never be removed from the league, since they provide a fast and easy way to increase play (compared to developing more sophisticated AI), make for interesting visual performances, and because all teams using them have a vested interest in keeping them in use. The latter points are also an important side-effect: large challenges cost a great deal of money to run, and are further intended to draw the attention of the popular and scientific communities, so there is a danger of flashy visual appearance taking precedence over a good evaluation.

While we have been focussing on examples in robotic soccer, similar issues occur in almost any robotic challenge, because of the nature and complexity of evaluating robotic systems. For example, recently the RoboCup rescue league (employing the NIST testbed) has moved to using angled floor panels (convex and concave) to replace flat surfaces in the simplest of its arenas. Rather than developing very general approaches to traversing terrain, teams are instead encouraged to deal with this precise type of angled flooring. While the NIST arena does graduate to more challenging areas that feature stepping fields that promote more diverse terrain, those restricting themselves to one area of the arena need only consider that specific solution. In the past, the RoboCup Rescue scoring has also demonstrated similar problems to the RoboCup Small-Sized soccer league, in promoting specialized hardware over general approaches. For example, a multiplied score could be tallied by detecting a single victim multiple times using different forms of sensing. This allowed a human operator to, for example, perceive a victim themselves through their robot's camera, and be assured of its position, and then simply use the alternative forms of sensing (e.g. a heat, motion, $CO_2$ detectors) provided on the robot to re-flag the location of the victim. Like the small-size league, this promotes a high score simply through equipping a robot with additional hardware that may not be necessary - a specialized solution - over the adaptivity and flexibility that we expect from good AI. The NIST arena has the additional problem of discouraging autonomous approaches significantly, by employing a scoring system that allows human-teleoperated robots to score far better than is possible for any current autonomous system. Since most teams are not fully autonomous, there is again a vested interest in keeping the scoring system as it is, as opposed to altering it to encourage better AI. While this still allows good comparisons between teleoperated solutions and a potential means of appropriately evaluating work in human-robot interaction, it limits the applicability of the challenge.

The remainder of this paper describes some of our work toward dealing the issues involved in competition-based robotic evaluation, which we see as responses to the problems of control and repeatability and those of focussing on narrow, specific solutions. To deal with issues of control and repeatability, we advocate the use of a mixed reality element in the challenge problem associated with a competition, along with a suitable breakdown in the core challenge to promote breadth. The next section describes our experiences with using this in a classroom environment from an evaluation standpoint. We then present an extension of the idea of breadth for appropriate competition-based evaluation through our design of the FIRA HuroCup rules (Baltes 2008), which are intended to encourage breadth and robustness to counter the narrow focus that typically occurs in competitions.

## Mixed Reality: Improving control and repeatability

Adding a mixed reality element to a competition challenge requires the creation of both physical and virtual elements for robots to perceive and act upon. Our approach to this is illustrated in Fig. 1. Robots (along with any other physical elements) are placed on a horizontally mounted LCD panel, on which can be projected any virtual elements desired by an application. An oblique view of the field is captured by a camera, and the image is interpolated and objects tracked via Ergo, our intelligent global vision software (Anderson & Baltes 2007b). Ergo reports the positions of any objects that have been defined for tracking via ethernet, and similarly a world server can report the positions of any objects on the virtual plane if they are not already being tracked by the camera itself (e.g. if their visual complexity precludes real-time camera tracking). Client programs controlling robots receive these descriptive messages, formulate commands for the robots they control, and communicate these via ethernet. This approach allows students developing client programs to be concerned with high-level object locations rather than low-level perception, and allows robots to perceive both physical and virtual elements, and interact with both.

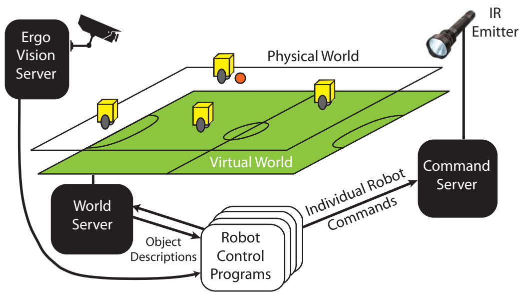


Figure 1: Overview of a mixed reality approach

Client program commands are received via a command server, and are batched for infrared transmission via IRDA. Any robot that can receive infrared signals can be used in this approach, as can any form of LCD, with the obvious constraint that the robot's physical size and the environment it inhabits must fit on the LCD. In practice, we have used everything from high-end Citizen Eco-Be micro robots to cheap infrared toys successfully, and 42" LCDs down to basic laptops. At an abstract level, the same approach is used in the RoboCup Mixed Reality Competition (which itself was adapted from our previous work with the E-League (Anderson *et al.* 2003)). In our work, all of the components shown in Fig. 1 are our own (and are available via our laboratory web site). Most of our components (e.g. the ability to have a camera offset rather than perfectly overhead) improve significantly on this basic model.

In prior work (Anderson & Baltes 2007a; 2007b) we have described the pedagogical approach used to apply this platform to undergraduate education. Here, we focus on its use in evaluation when employing a competition in such settings. We consistently employ competition-based evaluation in all our courses involving robotics, both to motivate students, and to allow them to see the and learn from the performance of different approaches, but continually had to deal with the problems of control and consistency in evaluation and fairness in comparison that have been described in the previous section.

From the standpoint of evaluation, adding a mixed reality element brings the advantages of both simulation and physical robotics together. Physical robots ensure all of the potential challenges faced by real robots exist, as well as the constraints imposed by the physical world. The ability to have physical objects other than robots also helps to ensure that the hard expectations of the real world are met (e.g. a physical ball can be used in playing soccer, removing any question of the physics of bouncing off a goalpost). The virtual elements, on the other hand, can be used to properly assert the control that is desired for evaluation, increasing repeatability from an experimental standpoint, and fairness in comparison from the standpoint of a competition. An example of this is shown in Fig. 2. This figure shows a team's approach to dynamic path planning being tested for the ability to avoid moving obstacles, on a 40" LCD using 2" remote controlled treaded tanks. The entrant must plan a path in real time while avoiding two randomly moving spots. The spots are virtual, and can be restarted at exactly the same places and move with the precise degree of randomness that a simulated world can guarantee across multiple trials (a virtual marker can also be used to more precisely place a robot as well). At the same time, the vision system itself can precisely track when the marker for a robot has occluded a spot, to a far better degree than a human observer.

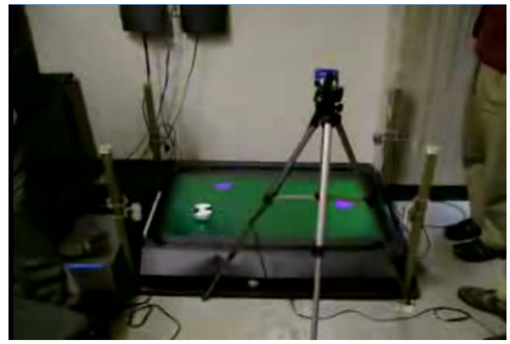

Figure 2: Perfectly repeatable moving obstacles in the virtual plane.

The automatic detection of such occlusions is an important part of consistency in evaluation brought by mixed reality environments, and the situation seen in this example adapts itself well to many other environments. The lower half of Fig. 3, for example, shows a virtual puck automatically being detected crossing the goal line in a mixed reality hockey game. Depending on what choices are made for virtual elements, this control and consistency can be introduced into any aspect of a mixed reality environment. In hockey, for example, stick-handling and manipulating a puck at high speeds is an important part of the game. Using a virtual puck in such an environment can allow more accurate modeling as to when puck control should be transferred from one agent to another than would be possible given the movement ability of very small wheeled robots.

The hockey application shown in the lower half of Fig. 3 also embodies a number of other concepts that are important for control in evaluation. There are times when we want the world to be more rich and challeng-


Figure 3: Mixed Reality Applications: above, Pac Man; below, Hockey.

ing than might be physically possible, and mixed reality nicely supports this. In hockey, for example, there are a number of types of shots, one obvious differentiation being those that send the puck sliding down the ice, vs. one that raises it in the air, over the top of players' sticks. Be defining these actions in terms of their effect on virtual elements, these actions can be included for a more realistic application than physical robots might support. In the hockey game depicted here, a series of virtual actions allows for these different effects on a virtual puck. From a consistency standpoint, these virtual actions are available to all competition entrants consistently, and are physically treated in the same manner, allowing the students to focus on the action and its consequences rather than on specialized hardware. It is equally possible to used a mixed reality environment to provide additional constraints that the physical environment itself cannot permit. For example, when dealing with the physics of moving in low-friction environments such as ice, the degree to which turns can be successfully made is much more strongly correlated to speed than is possible on an environment such as an LCD screen. Since the any client's choices of action are ultimately batched by the command server depicted in Fig. 1, the command server can have a veto over the success of any action on the physical robot, by failing to transmit that action or by transmitting a different outcome. Thus, depending on the speed of the robot, a turn might or might not be successful in the physics of the mixed reality application, even though it might in the physical world as given, allowing us even greater challenges than would otherwise be possible.

The wide possibilities for mixed reality applications, their potential differences in physics from the real world, and their similarities to gaming environments provide strong motivation to students, and the advantages for evaluation that are illustrated in hockey can be seen in any application. For example, the other application depicted in Fig. 3 is a mixed reality game of Pac Man, where robots occupy the roles of Pac Man and the ghosts associated with the game. Here, the spots that are consumed by Pac Man are virtual, and the robot's actions either succeed or fail precisely based on whether the robot is positioned to consume them. While others have implemented Pac Man physically using the Roomba (Dickenson *et al.* 2007), managing the physical aspect of this and resetting games so that spots to be consumed are properly placed for equal comparison is difficult.

While mixed reality brings with it the ability to have greater control over an evaluation, the problem of focussing on details and exploiting minutia in rules over developing good, robust AI systems remains. In order to use challenge-based evaluation in the classroom, the rules must encourage focussing on robustness and adaptability. While some of the elements of mixed reality (e.g. the effects of virtual actions being identical for all players) encourage focussing on using the tools intelligently, we structure challenges so that adaptability and robustness are a strong focus by linking a series of challenges together. This is also helpful from a scaffolding standpoint, so that students are gradually introduced to the complexities involved in embodied AI. Our first challenge is a racetrack environment, for example, requiring the ability to follow a path, while the second challenge involves crossing a soccer field with fixed obstacles and third involving moving obstacles. The latter two add the ability to plan paths as well as follow them, and dynamically alter paths respectively (the third is shown in Fig. 2. Each of these thus builds on those before it, as do the other challenges that follow. Students are informed at the outset that they may have to improve their existing approach in order to have a reasonable entry for the next challenge. They know only the very general outline of the next challenge, rather than the specific rules, and so they must anticipate that some changes from the current scenario may occur.

A sequence of appropriately-connected challenges can overcome the dynamic that leads to one-time, special purpose solutions. While our approach works well in the classroom, it must be adapted to a challenge environment for levels beyond an undergraduate classroom. For a research-oriented challenge environment, there is a difference in focus: work is at the cutting edge, as opposed to illustrating principles novel only to the students, and educational goals are secondary: there is no natural scaffolding to take advantage of, as there is in an educational setting. The challenge rules must be adaptable as technology advances to better meet the challenges posed, and the challenges are of necessity broader than those that would be adapted in the classroom. The environment itself also differs in that there is no top-down authority that would be natural to assume in a classroom environment: a committee typically organizes a challenge environment, usually selected from the participants. Others must see and understand the value of improving challenges over increasing their own likelihood of doing well in an evaluation. Adapting some of these ideas to such an environment in order to encourage breadth and robustness encompasses the remainder of this paper.

## HuroCup: Improving Evaluation through Breadth

To deal with the problem of emphasizing narrow special-purpose solutions over robust approaches that are adaptable to many situations, the organizers of robotics competitions must work hard to select challenges actively demand breadth, and scoring systems that punish attempts to rely solely on specialization. Participants must be made to see the value of the approach, and the complexity of achieving all this is increased in that keeping a broad range of challenges up-to-date technologically is much more difficult than doing so with one. As an example of what we feel is the way a competition must be designed to be effective evaluators for AI systems by promoting robustness and adaptability, this section describes our work in developing the current FIRA HuroCup competition for evaluating humanoid robots.

The FIRA HuroCup is the oldest humanoid robot competition, with the inaugural competition taking place in June 2002 with five teams. Since the HuroCup event is organized as part of the Federation of International Robosoccer Association, the initial plan was to develop a soccer competition for humanoid robots. However, it became quickly apparent that soccer did not provide a good benchmark problem for humanoid robots. Since soccer was played on a flat hardwood surface, many teams quickly developed efficient walking gaits and kicks for this surface. The main challenge then was to develop localization (where is the soccer player on the playing field?) and mapping (where are the other players and the ball?) methods for the players. However, localization and mapping are not specific problems for humanoid robots and research in these areas can be done without much change from wheeled or other walking robots.

Therefore, the HuroCup committee decided to focus on open research problems that are more closely associated with humanoid robots in particular. The main open research problems in humanoid robotics fall into several areas:

**Active Balancing** humanoid robots must be able to walk over various even and uneven surfaces. They also must be able to adapt their walk to changes in the weight and balance of the robot (Lift and Carry, Weight Lifting),

**Complex Motion Planning** humanoid robots can perform many different actions. The sheer number of these movements mean that they can not all be preprogrammed. Instead a humanoid robot must be able to plan new motions on-line (e.g. a new motion to lean over a barrier to operate a light switch or to pick up a box from under a table),

**Human-Robot Interaction** a humanoid robot must be able to interact naturally with a human which entails that it is able to understand speech, facial expressions, signs, and gestures as well as generate speech, facial expressions, and gestures.

Because one of the advantages of the humanoid form is its robustness and applicability to a wide variety of problems, some of these areas are naturally associated with robustness and breadth (e.g. walking vs. walking on uneven terrain vs. walking while carrying a load). Since this is a competition evaluating research, the researchers involved have a vested interest in leveraging this wide applicability, in their own research and to the public. Self interest along with many of the peripheral motivations of competitions discussed in Sec. can thus also be used as an advantage in encouraging breadth and robustness.

In deciding on challenge events, we and the other members of the HuroCup committee looked for those that would specifically advance research in these areas, as well as considering what would most encourage robust solutions and work well in a public challenge environment. To avoid exploiting rules in one large challenge environment attempting to encompass all humanoid skills, we instead focussed on dividing the FIRA HuroCup into a series of events that each test a subset of interacting humanoid skills. Scores in the individual challenges are summed, so that in order for an entry to do well in the HuroCup, a *single robot* must perform and score well across the range of events. Any special hardware development that focusses on doing well in one type of activity becomes redundant in others that do not require that specialization. Such additions can also be severely detrimental in two ways. First, given the limited time available in and around a competition, additional hardware and control that serves no purpose in some events draws support and resources away from the limited pool available to a team as a whole. More directly, the addition of such equipment may be strongly detrimental to the performance of other events (e.g. specialized arm motors making the robot more top-heavy, making adaptive balancing more difficult).

All HuroCup events require a fully autonomous robot that has all sensing and processing on board. No interaction from a human is allowed. HuroCup 2009 consists of the following eight events, some of which are depicted in Fig. 4:

**Sprint** the humanoid robot must walk a distance of 3.0m in a straight line forwards and then backwards.

Figure 4: Four events in the 2007 and 2008 HuroCup. From top: the Obstacle Run, Marathon, Basketball, and Lift-and-Carry.

This means that a robot must possess at least two fast walking gaits. This event is really intended as a starter event which allows beginning teams to score some points. The remaining events are more difficult.

**Obstacle Run** the humanoid robot must cross a 3.0m long region to reach the end zone without touching any of the obstacles. There are three types of obstacles: walls, holes, and gates. A robot must not step into a hole, but can crawl through a gate to reach the end zone.

**Penalty Kick** a humanoid robot tries to score against several goal keepers. This event is to include soccer related activities in HuroCup and is also considered relatively easy by most teams. In contrast to human soccer, the ball is placed randomly in an area in front of the robot.

**Lift and Carry** Lift and Carry was introduced in 2004. A robot must carry an increasing number of weights over an uneven stepping field. The stepping field is colour coded so that the robot can recognize steps. This is an advanced challenge and many teams have problems with it.

**Weight Lifting** The weight lifting competition was introduced to provide a slightly simpler active balancing challenge than Lift and Carry. A robot must lift as many CDs as possible. However, since we did not want to test the shoulder motor strength, the robot must walk 30cm with the weight low and then 30cm with the weight above its head. This means the centre of mass of the robot changes drastically, but predictably and the robot needs to compensate.

**Basketball** A humanoid robot must pick up a table tennis ball randomly placed in front of the robot and throw it into a basket.

**Marathon** A humanoid robot must cover a distance of 42.195m as fast as possible without being allowed to change its batteries. The event was the first HuroCup event that takes place out-doors, which means that teams must cope with more uneven surfaces and lighting conditions.

**Climbing Wall** a humanoid robot must climb up a wall where foot and hand holds were placed randomly. This is a new event in 2009.

The combination of events represent most of the range of activity expected of a humanoid, and the requirement of doing well in a range of events ensures breadth in evaluation. To do well in a dash, for example, a robot must have a fast start, but need not have fine control once it is moving. On the other hand, completing the marathon (Fig 4, second from top) requires following a lined track over a long period of time. Specialized hardware for either of these does not likely help the other. This is even more obviously seen in basketball, where teams have rarely attempted to use any special throwing motors, since the extra weight will tend to decrease performance in weight lifting and running events. Whereas most other robotics competitions still developed special technology to solve specific problems (e.g., teams in the small-sized league at RoboCup have developed special rubber mixtures to increase maximum acceleration of their robots and a range of kicking devices), HuroCup robots are still a core humanoid robot with two legs, two arms, and a camera.

The events are constantly updated to reflect the current state of the art. In 2009, for example, the climbing wall event was added, but major changes to other events were also introduced. In the lift and carry, the robot must pick up the weightlifting bar whereas previously, it could start with the bar in hand. The distances for the sprint and obstacle run were increased from the 2008 distance of 1.2m to 3.0m. The stepping obstacle was replaced by a hole/pit in the obstacle run, and the marathon will now be performed outdoors.

We believe that properly steering the evolution of a competition can go a long way in removing the focus on narrow, specialized solutions. To consider an example from the HuroCup, in 2004, about half of the teams used infra-red distance sensors to detect the obstacles in the obstacle run - an obvious piece of specialized hardware that does not reflect how humans deal with this problem intelligently. There was a heated discussion whether these sensors should be made illegal. The organizers felt that more human-like sensing was desirable, but instead of simply disallowing those types of sensors, the obstacle run competition was extended to include gate and hole obstacles, which cannot be detected easily by infra-red distance sensors. This meant that infra-red distance sensors were not the must-have sensor to do well in the obstacle run anymore. This led to fewer teams using infra-red distance sensors without having to disallow them in the rules and few teams complained when infra-red sensors were finally disallowed for 2009.

The organization of the HuroCup competition also incorporates lessons learned from other robotics competitions, to make improvements from both participant and spectator viewpoints. Teams rotate swiftly through the events (e.g. every one to three minutes a new robot will perform a task), and thus spectators do not need to watch a badly performing robot for 30 minutes. All events are repeated on different days, allowing a second chance when technical problems occur. This means that the HuroCup competition *always* starts on time and does not suffer the problem of long delays, because teams claim that technical difficulties are preventing their only opportunity. It also means HuroCup teams have many different opportunities to show and discuss their robotics work with spectators and other teams. A standard HuroCup event runs over four days with four to six events per day (roughly every two hours). This is far more than the six to eight games a humanoid soccer team plays in a competition and it means that teams also need to be concerned about the robustness of their robots to a greater degree.

## Conclusion

To summarize, robotic challenges are in part a reaction to the difficulty of repeatability and controlled experimentation in complex robotic domains. The difficulty with these robotic challenges is that the rules must be very precise in order to be able to appropriately compare solutions and limit approaches, while these same very precise rules discourage the development of robust, generally-applicable approaches. In this paper, we have shown techniques that can help deal with the issue of control and repeatability, through the use of a mixed reality approach. While this alone does not encourage breadth and robustness in solutions, we have described how we employ this in a classroom setting to deal with this problem, as well as how to apply such techniques on a much larger and broader scale, using our work with the FIRA HuroCup as an example.

We believe that the key to improving robotics competitions is to include technological advances that can help improve control, such as those described here, and in devising rules and scoring that specifically demand and test for breadth and adaptability, while forcing autonomous processing. The approach taken in the HuroCup, and the continued evolution of this competition, we believe, serves as a model for competition-based evaluation that stresses breadth and robustness over highly-specialized solutions.

## References


Anderson, J., and Baltes, J. 2007a. A mixed reality approach to undergraduate robotics education. In *Proceedings of AAAI-07 (Robot Exhibition Papers)*.

Anderson, J., and Baltes, J. 2007b. A pragmatic global vision system for educational robotics. In *Proceedings of the AAAI Spring Symp. on Robots and Robot Venues: Resources for AI Education*, 1–6.

Anderson, J.; Baltes, J.; Livingston, D.; Sklar, E.; and Tower, J. 2003. Toward an undergraduate league for RoboCup. In *Proceedings of RoboCup-2003*.

Baltes, J. 2008. HuroCup competition rules. http://www.fira.net/soccer/hurosot/overview.html.

Dickenson, B.; Jenkins, C.; Mosely, M.; Bloom, D.; and Hartmann, D. 2007. Roomba pac-man: Teaching autonomous robotics through embodied gaming. In *Proceedings of the AAAI Spring Symp. on Robots and Robot Venues: Resources for AI Education*, 35–39.

Fox, S. 2007. Finding the right robot competition: Targeting non-engineering undergraduates. In *Proceedings of the AAAI Spring Symp. on Robots and Robot Venues: Resources for AI Education*, 49–52.

Gerkey, B.; Vaughan, R.; and Howard, A. 2003. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*, 317–323.

Hamner, E.; Lauwers, T.; Bernstein, D.; Nourbakhsh, I.; and DiSalvo, C. 2008. Robot diaries: Broadening participation in the computer science pipeline through social technical exploration. In *Proceedings of the AAAI Spring Symp. on Using AI to Motivate Greater Participation in Computer Science*, 38–43.

Jacoff, A.; Messina, E.; Weiss, B. A.; Tadokoro, S.; and Nakagawa, Y. 2003. Test arenas and performance metrics for urban search and rescue robots. In *Proceedings of IROS-2003*, 3396–3403.