CHAPTER 5

Gaseous interactions
The paper discussed in this chapter is about a home robotics competition called RoboCup@Home. The hominid writer of this dissertation is also the co-founder and executive of this competition. At the last competition (2009, in Graz, Austria) there were more than 200 scientists in more than 20 teams coming from more than 14 countries from around the planet and in 2010 (in Singapore) more than 25 teams are expected. The @Home competition is part of the RoboCup Federation, an organization that organizes robotic competitions and symposia for more than a decade. This year (2009) there were more than 440 teams from more than 60 countries in total, adding up to several thousands of scientists participating in total. This chapter reports on the progress and the way to measure the progress of RoboCup@Home. The critiques on the original article from peer reviewers was rather severe, because the proposed benchmarking system fluctuates over time, depending on the results from the competitions of previous years. The reason for the yearly changes is that the executives of the @Home competitions are adjusting the sorting machine because it is unknown which sorting machine will lead to intelligent robots. Structuralists have difficulties with this kind of thinking and require a method or data set that remains fixed over time. The disadvantage of the structuralistic methods of thinking is that scientists create ITO solutions that optimize only a very local problem. From a post-structuralistic point of view the fluctuating principles from RoboCup@Home could imply that there is a singularity in the neighborhood. In the literature from non-linear dynamical systems fluctuations usually precede bifurcations (Prigogine (1984)).

This chapter describes an environment where GAI can be tested. It includes a rich set of physical and mental stimuli and poses international benchmarks to compare the different approaches available. An environment as in RoboCup@Home is a typical place where long running experiments can and should be set up. Another reason to include this chapter is to demonstrate that the general part of the way of thinking as set out in the beginning of the book is not only derived from general principles found in physics, biology and neurodevelopment. It can also be applied to many other fields of research, most of which are unknown to the author. This one is known to the author and serves as an example of the power of thinking in a generative manner.

This chapter is interesting for several reasons. The most important one is that due to the complexity of the tasks (involving robots and humans) the competition pushes teams toward finding bifurcations in their thinking and in the methodologies they use to create ever more intelligent machines. Traditional methods seem to work only partially in @Home and the participants (both humans and robots) cannot rely (completely)
on traditional methods. This stimulation of creativity has resulted, in just a couple of years, in an increment in the speed of research of the research area of domestic service robots. The teams generate many solutions (both in hardware and in software). The competition functions as a sorting mechanism to select only the most reliable and the most useful functionalities of the intelligent machines.

Another important aspect is that ROBOCUP@HOME is actively creating measurement tools to steer the developing processes of intelligent machines. These tools are never based on the internal mechanisms of the machines, but always on (relatively) easy to inspect aspects of the robots, such as behaviors and linguistic expressions. In ROBOCUP@HOME it is important that for any human who is interacting with the intelligent machines the judgment of the performance of and interactions with these machines should be intuitive.

Although it is difficult to prove, there are probably already phase transitions going on in ROBOCUP@HOME. In just four years time the best performance went from dragging a beer-crate with a rather large hook to real-time machine learning where the robot is able to generalize on objects (a sorting mechanism) and to copy behaviors of hominids (a behavior generator).

The next article has been accepted for the journal of Interaction Studies in 2009 under the title “Robocup@home: Developing and benchmarking domestic service robots through scientific competitions”. The article is a follow up of “Robocup x: A proposal for a new league where robocup goes real world” (van der Zant and Wisspeintner (2005)) and “Robotic Soccer, chapter RoboCup@Home: Creating and Benchmarking Tomorrows Service Robot Applications” (van der Zant and Wisspeintner (2007)), combined with the latest insights from the humans organizing the competition.
5.1 Introduction

Being part of the RoboCup initiative, the ROBOCUP@HOME league targets the development and deployment of autonomous service and assistive robot technology being essential for future personal domestic applications. The domain of domestic service and assistive robotics implicates a wide range of possible problems. The primary reasons for this include the large amount of uncertainty in the dynamic and non-standardized environments of the real world, and the related human interaction. Furthermore, the application orientation requires a large effort towards high level integration combined with a demand for general robustness of the systems. This chapter details the need for interdisciplinary community effort to iteratively identify related problems, to define benchmarks, to test and, finally, to solve the problems. The concepts and the implementation of the ROBOCUP@HOME initiative as a combination of scientific exchange and competition is presented as an efficient method to accelerate and direct technological and scientific progress in the domain of domestic service robots. Finally, the progress in terms of performance increase in the benchmarks and technological advancements is evaluated and discussed.
5.1. Introduction

5.1.1 Background

The general idea of personal Domestic Service Robotics (DSR) has been around for a long time, but it is a comparably young research topic. The aim of creating useful, autonomous, multipurpose personal assistant robots which can interact with humans and objects in the real world in a natural way poses a large number of unsolved problems across many scientific disciplines.

There have been many successful and impressive demonstrations of robot technology in the past. In DSR, one focus-and one of the main difficulties-is the interaction with the real world, instead of operating under constrained settings and strictly defined environmental conditions as opposed to e.g. industrial robotics. These systems must cope with a large amount of uncertainty. A natural home environment, for example, is not specified in size, shape, appearance, the kind of objects contained in it, lighting and acoustic conditions, the kind and number of residents, etc. Furthermore, as objects and people can move, disappear and reappear, the environment is dynamic. The system must be able to manipulate objects in various locations and from different heights, and it needs to be capable of locomotion on different terrains. When interacting with humans, the system should possess some basic (social) intelligence and should be able to distinguish different people. Last but not least, safe and robust operation of these systems in such uncertain and dynamic environments is a fundamental requirement for their future acceptance and general applicability.

The dynamical and uncertain environment of ROBOCUP@HOME where autonomous intelligent machines have to perform complex tasks are the perfect place to test GAI. This has not occurred yet in ROBOCUP@HOME due to the fact that the theory of GAI is too young to have been put to the test here. In fact, ROBOCUP@HOME is not the alone in its efforts to dynamically stimulate the progress of AI and Robotics. The total effort of the RoboCup Federation can be seen as a iterative process towards for fostering these kind of technologies. ROBOCUP@HOME is pushing this entire process to the limits though. It has change incorporated into its fundamental structures and is in constant flux. The environment of ROBOCUP@HOME is the least specified and it is the only league that also uses the real world for the competition, instead of only a scenario such as a soccer field. It can be argued that of all the leagues in the RoboCup Federation ROBOCUP@HOME is progressing the fastest.

The creation of such autonomous systems requires the integration of a large set of
abilities and technologies. Examples include human-robot interaction (speech, gesture, person, face recognition and tracking, among others), navigation and mapping, reasoning, planning, behavior control, object recognition, object manipulation or tracking of objects. With regard to artificial intelligence, the systems should contain adaptive but robust behavior and planning methods, social intelligence, and learning capabilities. Intuitive programming methods (instead of entering computer code) are required for a broad acceptance and usability. Appropriate procedures should, for instance, enable the robot operator to teach new behaviors and environments via voice or gesture commands. As future households will most likely contain more intelligent electronic devices capable of communicating with each other, ambient intelligence, including the use of the Internet as a common knowledge base, will certainly play a more important role.

Just very recently, progress in these research fields, as well as progress and standardization in related hardware and software development, has led to an increase in availability of required methods and components for DSR. This includes the availability of software frameworks for robot control (e.g. Carmen\(^1\) (Montemerlo et al., 2003), Player/Stage\(^2\) (Gerkey et al., 2003), MRPT\(^3\), MRS\(^4\)), simulation (e.g. USARSim\(^5\) (Balakirsky, 2006)), and open source software libraries containing algorithms for computer vision (e.g. OpenCV\(^6\) with diverse applications as shown in (Bradski and Pisaresvky, 2000)) or robot control (e.g. Orocos\(^7\) (Bruyninckx, 2001)). On the hardware side, robot construction kits (e.g. VolksBot\(^8\) (Wisspeintner and Novak, 2007) and base platforms

---

\(^1\)Carnegie Mellon Robot Navigation Toolkit (http://carmen.sourceforge.net/)
\(^2\)The Player/Stage Project (http://playerstage.sourceforge.net/)
\(^3\)The Mobile Robot Programming Toolkit (http://babel.isa.uma.es/mrpt/index.php/Main_Page)
\(^4\)Microsoft Robotics Studio (http://msdn.microsoft.com/en-us/robotics/default.aspx)
\(^5\)Unified System for Automation and Robot Simulation (http://sourceforge.net/projects/usarsim)
\(^6\)The Open Computer Vision Library (http://sourceforge.net/projects/opencv/)
\(^7\)Open Robot Control Software (http://www.orocos.org/)
\(^8\)VolksBot Robot (http://www.volksbot.de/index-en.php)
(e.g. ActivRobots\textsuperscript{1}), faster and energy efficient computation or light weight manipulation devices (e.g. Katana\textsuperscript{2}) as well as miniature sensors (e.g LIDAR\textsuperscript{3}) are available.

In sum, increased availability, accessibility and compatibility of these essential robot components enables research groups not only to address a small subset of the mentioned above challenges in DSR, but also to address the problem as a whole. Obviously, DSR is not solely about integrating existing solutions. But the consequent reuse of existing technology can help to save time and effort, so researchers can focus on a particular research field while maintaining a fully operable robot platform.

This is also confirmed by the presence of some rather specialized service robotic applications on the market. Such applications include floor cleaning (e.g Roomba and Scooba\textsuperscript{4}), lawn mowing (e.g. Robomow\textsuperscript{5}) and surveillance (e.g. Robowatch\textsuperscript{6}). Still, these service robots do not possess the properties of a multipurpose autonomous and intelligent domestic service robot.

To state it differently: the robots on the market are ITO solutions. These robots are optimized for a single or a few tasks. They exhibit no form of generalization capabilities or anything that most humans would classify as intelligent behavior. On the other hand having all sorts of structure generating pieces of software available it is easier to create networks of processes. For this it is essential that machine learning technologies are applied on top of the processes already available in the robots. New relations between internal states and external information has to be found by the software, not only by the humans who program it. This requires that the robots operate for long periods of time so that the affordances can be found.

Prominent examples of domestic and personal assistant robot research projects include ReadyBot\textsuperscript{7}, and PR2\textsuperscript{8}. Wakamaru \textsuperscript{9} and PaPeRo\textsuperscript{10} focus more on social interaction studies. Many of these projects address relevant aspects of DSR. Still, what appears to be missing is a joint, continuous international and multidisciplinary research

\begin{footnotesize}

\begin{itemize}
\item \textsuperscript{1}ActivRobots (http://www.activrobots.com)
\item \textsuperscript{2}Katana robotic arm (http://www.neuronics.ch/)
\item \textsuperscript{3}Hokuyo Sensor Technology (http://www.hokuyo-aut.jp/)
\item \textsuperscript{4}iRobot (http://irobot.com)
\item \textsuperscript{5}Robomow (http://www.friendlyrobotics.com)
\item \textsuperscript{6}Robowatch (http://robowatch.com)
\item \textsuperscript{7}ReadyBot (http://www.readybot.com/)
\item \textsuperscript{8}PR2 (http://www.willowgarage.com/)
\item \textsuperscript{9}Wakamaru (http://www.mhi.co.jp/kobe/wakamaru/english/)
\item \textsuperscript{10}PaPeRo (http://www.nec.co.jp/robot/english/robotcenter_e.html)
\end{itemize}
\end{footnotesize}
and development effort which also includes the aspect of application-oriented benchmarking of systems in DSR.

With this motivation, the authors initiated the ROBOCUP@HOME competitions in 2005 (van der Zant and Wisspeintner, 2005, 2007). The ROBOCUP@HOME league targets the development and deployment of autonomous service and assistive robot technology as being essential for future personal domestic applications. It is part of the international RoboCup initiative, and it is the largest annual service and home robotic competition worldwide. The ROBOCUP@HOME tournaments are organized in independent test scenarios, which are used to benchmark the robots’ abilities and performance in a realistic non-standardized home environment. More specifically, ROBOCUP@HOME aims to proffer a combination of interdisciplinary community building, scientific exchange and competition, which iteratively defines benchmarks and performance metrics on which service robots can be evaluated and compared in a realistic domestic environment.

Since the real world is not standardized, measuring the performance of non standardized robots acting in it is a difficult task. The experimental paradigm to evaluate the complex robotic systems has to use consequent scientific analysis to improve on itself. Measuring the performance of the robots requires continuous reconsideration of the methodologies used since both the robots (their capabilities) and their operation environment (and the robot’s tasks) will definitively change over time. This co-evolutionary development process, the feedback and refinement procedure, is a key element of the ROBOCUP@HOME league. In our case, the tools are specific benchmarks which test certain robot abilities and the measurement of the robots’ performance.

ROBOCUP@HOME also measures, in a scientific and quantifiable manner, the performance of complex systems. We firmly believe that creating and applying this experimental paradigm can greatly improve robotics developments.
ROBOCUP@HOME functions as a sorting machine on technologies generated by research groups. The generative capabilities of the rather large amount of scientists together with a common sorting machine is really something different than having many individual research groups with no or completely different sorting machines. It is the difference between a world-wide coordinated research programme or many small programmes with little or no interaction. Research programmes that were already in progress form interactions with each other, enhancing the generative capabilities of the entire group of researchers and their research projects. In other words, the research teams (generators) form a network (meshwork) where, in an iterative manner, the research progresses. The interactions between research groups and the cyclic flux of the research projects combined with the competition (sorting machine) form an abstract search machine on its way to the next singularity.

This article thus addresses the problem of benchmarking DSR through scientific competitions by presenting the approach followed in the ROBOCUP@HOME initiative. The article contains several contributions:

- it presents an overview of benchmarking through competitions, describing other existing competitions and highlighting the unique features of ROBOCUP@HOME;

- it describes the underlying concept of the @Home competition and its implementation into a framework for benchmarking in DSR which aims to be a common testbed for application development;

- it provides a detailed analysis of the results from different viewpoints that are of crucial importance for assessing the actual performance of DSR and for planning future tests and other competitions.

The remainder of the article is organized as follows: The next section gives an overview of the state of the art in robotic benchmarking and DSR. Then, the concept and the implementation of the @HOME competition are presented. Section 5.4 will evaluate the benchmarking results of the past several years and discuss the observed increase in performance, the scientific achievements and the importance of a vital community. The article concludes with an outlook on short and mid-term goals.
5.2 Benchmarking Domestic Service Robotics

Benchmarking has been recognized as a fundamental activity to advance robotic technology (del Pobil, 2006; Sabanovic et al., 2006) and many activities are in progress. Some projects and special groups are working on defining standard benchmarking methodologies and data sets for many robotic problems, like Human-Robot interaction (HRI), SLAM, or navigation. Examples for such initiatives are the EURON Benchmarking Initiative\(^1\), the international workshops on Benchmarks in Robotics Research and on Performance Evaluation and Benchmarking for Intelligent Robots and Systems, held since 2006\(^2\), the Rawseeds project\(^3\), which aims to create standard benchmarks especially for localization and mapping, and the RoSta project\(^4\), which focuses on standardization and reference architectures.

Benchmarking can be distinguished in two classes: system benchmarking, where the robotic system is evaluated as a whole, and component benchmarking, where single functionality is evaluated. Component benchmarking is integral for comparing different solutions to a specific problem and for identifying the best algorithms and approaches. Among the many examples, much effort has been put on mapping and SLAM (e.g. (Howard and Roy, 2003; Fontana et al., 2008)), and navigation (e.g. (Baltes, 2000; Munoz et al., 2007; Calisi et al., 2008)). While component benchmarking is useful for directly comparing different techniques of solving a specific problem, it is not sufficient for assessing the general performance of a robot with respect to a class of applications. Indeed, the best solution for a specific problem may be unfeasible or inconvenient when integrated with other components that compose a robotic application. On the other hand, system benchmarking offers an effective way to measure the performance of an entire robotic system in the accomplishment of complex tasks, as such tasks require the cooperation of various sub-systems or approaches.

In this kind of benchmarking, a standard reference environment, reference tasks and related performance metrics are to be defined. Examples of system benchmarking are given in the fields of interactive robots (Kahn et al., 2007) and of socially assistive robots (Feil-Seifer et al., 2007).

---

\(^1\)http://www.euron.org/activities/benchmarks/index
\(^2\)All these workshops are summarized in http://www.robot.uji.es/EURON/en/index.htm.
\(^3\)http://www.rawseeds.org/
\(^4\)http://www.robot-standards.eu/
ROBOCUP@HOME clearly operate on the system benchmarking level. The components in GAI are flexible. For the testing of intelligent robots the complete system has to be tested, in many situation with many different kind of tasks. RoboCup in general functions as sorting machines on the technologies generated by the scientific teams that participate. It is essential that the individual leagues of the RoboCup federation keep changing their rules and reset their goals in order to avoid local optima out of which it might be difficult to escape.

When defining standard benchmarks, two common problems arise:

- The difficulty of defining a benchmark that is commonly accepted by the community (this is due to differing viewpoints on a problem from separate research groups);
- The risk of creating specialized solutions for a certain benchmark or problem that cannot be applied in real world applications.

To avoid these problems, scientific competitions have proven to be a very adequate method because:

- benchmarks are usually discussed and then accepted by all the participants;
- participants are usually required to solve multiple benchmarks. These benchmarks vary over the years, thus providing for a disadvantage in using solutions that are too specialized.

Moreover, competitions provide an effective means of interaction and communication among research groups because they are often associated with scientific conferences or workshops and provide participants a large audience for their research efforts. Finally, annual competitions provide regular feedback on performance increases and allow for establishing medium-term projects.

Among the many robotic competitions, the AAAI Mobile Robot Competitions were one of the first, being established in 1992 (Balch and Yanco, 2002). RoboCup (founded in 1997) (Kitano et al., 1997) currently has the largest amount of participants (e.g. 440 teams with more than 2,600 participants from 35 countries in 2006). The DARPA Grand Challenge is probably the most recognized in terms of public and media attention and the one that is most directly application-oriented.
Furthermore, educational contests, such as EUROBOT\(^1\) or RoboCup Junior\(^2\), are organized with the main goal of presenting robotics to young students. Thus, they deal with simpler tasks and robotic platforms.

All of these competitions have obtained very relevant results, which are analyzed in the following:

**AAAI**  
AAAI Mobile Robot Competitions are held in conjunction with the AAAI and (sometimes) IJCAI Conferences on Artificial Intelligence. Thus, it offers great visibility within the AI scientific community. Many important scientific and technological achievements demonstrated during these competitions have been reported (Balch and Yanco, 2002). Although these competitions offer a relevant suite for benchmarking AI and robotics technology with relevance for real-life applications, their focus and benchmarks change heavily on a yearly basis. This change of focus makes it difficult to approach the problems in a continuous and iterative way and to build up a community with a long-term goal.

Although the ideas behind this competition are relevant for GAI and it is positive that the competition is in constant flux, it might be that there is too much change from a year to year basis to scaffold enough mechanisms that can be reused. The lack of a long-term goal means that the constructed AI is not directed and has little foundations to build its progress upon.

**RoboCup Soccer**  
RoboCup soccer competitions have set their sights on soccer because it is the most common sport worldwide. Its long-term goal is to compete with the human soccer world champion team by 2050. Moreover, as opposed to AAAI competitions, RoboCup events put the main focus on the competition and offer the possibility to discuss scientific achievements in a small and more focused RoboCup Symposium. RoboCup has proven to provide an efficient means of interaction and communication among research groups. It combines scientific research, competition, benchmarking and reality checks on various concepts. Performance is measured on a yearly basis. However, having a specific focus on soccer also presents some limitations. The main limitation is an over-specialization of solutions due to more or less fixed environmental

---

\(^1\)http://www.eurobot.org/
\(^2\)http://rcj.sci.brooklyn.cuny.edu/
conditions and rules. For example, in the middle-size league, where the design of the robot is a major issue, all the teams rapidly converged toward the same hardware architecture (catadioptric cameras and omni-directional driving robots) which was highly optimized on the provided scenario. Although this causes an immediate improvement of the average performance in the competition, it contains the danger of running into a single suboptimal solution which can not be applied to a real-world setting.

The soccer competitions of RoboCup could be seen as a single system. Instead of changing much of the rules within a sub-league (they do change but slowly) new leagues are created more or less continuously with a different focus. This creative mechanism can be regarded as the 'change' factor which is built into the ROBOCUP-@HOME league. Creating new leagues also implies letting go of leagues which have proven their worth but are no longer changing. It is important that the meshworks of scientists and technologies are transferred from league to league.

Robot Rescue Another example is given by Search and Rescue Robotics. Rescue competitions started in 2000 within the AAAI Mobile Robot Competition (Meeden et al., 2000) and since 2001 within the RoboCup Rescue initiative (Kitano and Tadokoro, 2001). RoboCup Rescue competitions have defined standard rescue arenas and tasks for benchmarking robotic search and rescue missions and for measuring an increase in performance of the rescue robotic technology in a standardized abstracted environment. Within the Rescue competitions, common metrics for HRI have been defined (Steinfeld et al., 2006) and effective evaluation of HRI techniques have been carried out (Yanco et al., 2004; Drury et al., 2005), with a specific focus on the interfaces used by operators to interact remotely with the rescue robots. Nevertheless, this indirect kind of HRI via an operator station involving semi-autonomy and remote control is different to what is required in most DSR tasks, where the focus is a direct and more natural interaction and full autonomy. Still, one can think of certain DSR applications where such kind interaction is desired, e.g. to monitor and communicate with nursing cases remotely. Therefore, the concepts of the rescue robot initiative with respect to benchmarking, as well as rescue robot technology, are highly relevant when defining benchmarks for DSR.
In the rescue competitions of RoboCup the environment has more uncertainties than in soccer and the complexity of the tasks increases year by year. It is interesting that this league does not need to have many sub-leagues, as in the soccer case, but incorporates the idea of change. The league uses well defined measures to structure the change. It is clear that the initiators of ROBOCUP@HOME took a close look at RoboCup Rescue while defining that competition. In ROBOCUP@HOME the idea of change is pushed even further, while also using metrics to steer the generative processes.

**DARPA Grand Challenge** The DARPA Grand Challenges\(^1\) were the most prominent robotic competitions to date. GPS navigation together with multimodal sensor data fusion were commonly used to face the uncertainties and dynamics of real-world application scenarios. In the Urban Challenge, even real traffic rules were applied, but the complexity was limited by simplifying the cognition tasks. Contextual information was entered into a predefined map, the route network definition file (RNDF), which consisted of way points with GPS coordinates, connection types, traffic signs, number of lanes, width of lanes, etc. Participating in these challenges required a lot of effort, as the joint work of different research groups and industries with complementary competencies was a critical factor. At this time, it is uncertain if this initiative will be continued in the future and to what capacity.

**Service Robotic Competitions** Initiatives that are directly related to Domestic Service Robotics mainly aim at a single specific task. For example, the AHRC Vacuum Contest\(^2\) and the 2002 IROS Cleaning Contest\(^3\) (Prassler, Erwin and Hägele, Martin and Siegwart, Roland, 2006) are focused only on floor cleaning, while ROBOEXOTICA\(^4\) concentrates only on robots preparing and serving cocktails. A more general initiative is given by the ICRA HRI Challenge\(^5\), and it is motivated by the fact that “the effectiveness of a robot engaging in HRI must be evaluated by human users who got the chance to interact with the robot for a sufficiently long period of time.” However, it is still in preliminary stages because evaluation criteria for benchmarking the performance have not been defined.

---

\(^1\)http://www.darpa.mil/grandchallenge/index.asp
\(^2\)http://www.botlanta.org/
\(^3\)http://robotika.cz/competitions/cleaning2002/en
\(^4\)http://www.roboexotica.org/en/mainentry.htm
\(^5\)http://lasa.epfl.ch/icra08/hric.php
Although these initiatives are very relevant to the field, it is evident that there is no major international annual competition in the field of Domestic Service Robotics that can be considered a continuous integrated system benchmarking activity.

As can be seen in all the robotics competitions, defining a good sorting machine (competition) is not a straightforward task. Especially with robotic machines with a large amount of complexity the creation of the correct sorting mechanisms is a research programme in itself. It entails more than collecting a data set for benchmarking or defining a (set of) task(s). The sorting mechanisms have to evolve together with the solutions generated by the participants. The evolution of the sorting machine is not uncommon. In biology it is a well established fact that the criteria for selection co-evolve with the species.

Following the successful lines of RoboCup competitions and the experiences amassed at other competitions related to DSR, we have set up the ROBOCUP@HOME initiative, which is a system benchmarking activity for domestic service robotics implemented as an annual competition (van der Zant and Wisspeintner, 2005, 2007).

More specifically, ROBOCUP@HOME is a combination of scientific exchange and competition that provides standard benchmarks (called ”Tests”) and performance metrics on which personal domestic service robots can be evaluated and compared in a realistic domestic environment.

5.3 The @HOME initiative

With the insights and experiences gained from participating in and from observing and organizing various robotic competitions, we started with the conceptual design of a new competition for DSR in 2005.

5.3.1 Concept

The following considerations and criteria act as the basis of a common agreement for the ROBOCUP@HOME initiative.
Uncertainty  To reflect the uncertainty immanent in every real-world setting, the rules should not specify or limit any more qualities of a task than (absolutely) necessary. This complies with the requirement of providing a lean set of rules. Moreover, it encourages robust solutions that remain functional over a wide number of particular situations under as many circumstances as possible. This way, object positions or environmental characteristics, such as lighting conditions, are not specified and the scenario setup is changed frequently.

Or one could also state that the sorting machine should not sort out promising or fit solutions. A too strictly specified sorting mechanism steers toward brittle solutions and local optimization procedures. Locality is exactly what is to be avoided if the aim is the creation of intelligent machines.

Extendable framework for benchmarking  With the aim of benchmarking myriad robot capabilities for DSR with many of them yet to be developed, the framework for the competition needs to allow for constant evolution and modular enhancement of itself. The framework will consist of an initial set of independent benchmarks (called “tests”) all testing an individual set of relevant capabilities in DSR. Over time, when an increase in performance in the individual tests is observed, these tests are either enhanced by making the tasks more difficult or the tasks are merged together to form a more integrated, and therefore more realistic, application scenario.

The competition has to be open-ended to avoid local optima.

Autonomy  Robots in the @HOME league are required to be fully autonomous and mobile. That is, robots must complete benchmarks without being controlled remotely. To lower the demands for on-board computers, external computation is permitted as long as nobody interacts with the external computer during the test. To foster demonstration and use of new approaches, external devices the robot can interact with (external cameras, sensors, etc.) are allowed in certain tests. It may appear that instructions given by a human acting with the robot are a kind of tele-operation, but the execution of a given high-level task such as “Bring me object A” incorporates autonomy in terms of task decomposition, decision making, perception, task planning, and task execution. We enforce this autonomy by the uncertainty inherent in the environment. Open-loop
control of the robots should be hindered from the beginning by applying randomness in the start procedure (i.e. the robot entering the scenario on its own) and having strict time limits for the setup of the robot.

Autonomy is essential for the automated construction of intelligent systems. Interaction with the world alone is a slow process to create intelligence as has been demonstrated by natural evolution. The same process also demonstrated learning mechanisms within an individual in combination of training by peers can be a fast process to tune the animal for its surroundings. This implies that human-robot interaction of human-machine interaction is an essential ingredient for training AI.

**Natural interaction** In order to inhibit control of a robot by keyboard commands, currently, interaction with a robot be natural in all tests. This means that the interaction is either done via natural language or gestures (no keyboard control). Other modes of interaction like the use of touch screens or advanced remote controls can and should be demonstrated in the form of technical challenges (see Open Challenge, Demo Challenge and Finals in Section 5.3.3), where these restrictions are not applied. Then, corresponding solutions are to be integrated and allowed in future competitions. Moreover, haptic interaction (touching the robot) should further foster development of intuitive modes of control and interaction (instead of using a standard computer keyboard) and consider future use by the target audience: the general public, laymen in robotics, or elderly and disabled people.

The natural interaction allows for the grounding of the robots through steering, instead of programming locally optimized solutions. If robotic systems are trained for a long period of time it is mandatory to structure the training programs of robots (the sorting machines), not unlike the structuring of the educational system of humans.

**Benchmarking in uncertain conditions** The home environment in which the benchmarks take place is not standardized: It contains natural, non-standardized objects and varies over the years. Examples of previous competition environments are given in Figure 5.1. The degree of uncertainty contained in the benchmarks is high as the environment is hardly specified in size, shape, contained objects, kind of walls or floor, lighting and sound conditions, etc. Especially the interaction with humans adds to this uncertainty. Therefore, one of the challenges when defining a benchmark is to maintain
approximately the same difficulty for all participants while boundary conditions may change, so the performance can be compared. By increasing the amount of uncertainty over time, robust solutions are expected to emerge.

A changing environment requires adaptive mechanisms (generators) to cope with it. Static environments cause optimization to a local optimum. This implies that intelligence systems can only arise in dynamic environments in which new situations arises continuously.

**Fostering a wide range of approaches and solutions** The rules should be kept as unrestrictive as possible, and the benchmarks are to be defined in such way that the solutions for the given problems are not implicit. This approach requires a high level of common sense and agreement from the teams and the community, as trivial and undesired solutions to certain problems can not be completely avoided. Also, participants should have the choice to select certain benchmarks according to their background, skills, and their robot’s capabilities. Besides having predefined benchmarks the teams can select from, the competition will also offer the possibility of demonstrating new abilities and scientific results or applications not yet covered in the tests. These new aspects can later be used to enhance the benchmarks in the future.

**Multidisciplinary community** Putting few restrictions on the robots participating and providing the freedom to select benchmarks and approaches should motivate teams from different research backgrounds to participate in and to contribute to a growing community, one which fosters the exchange of multidisciplinary scientific and technological knowledge. Furthermore, the development of a common vision about the goals, as well as common sense and fair play in the competition, are required. Feedback from the teams is further needed to iteratively develop the competition.

**Generating public awareness** The competition should also generate interest from a non-technical, public audience by demonstrating usefulness in daily life, future applications and social relevance. This way public awareness for DSR should be increased, and links to the industry should be established.
Figure 5.1: From top to bottom: The @HOME scenario in 2006, 2007 and 2008
5.3.2 Defining key features

Before starting with the implementation of the benchmarks, an initial set of robot key features (abilities and properties) was derived from an analysis of the state of the art in DSR and our experiences and observations from other robotic competitions. These features help to design the benchmarks and the score system for the competition. Furthermore, these features allow for a later analysis of the teams’ performances and help to develop and later enhance the competition in a structured way. As the competition and its benchmarks are expected to evolve over time, the key features and their weights in the competition are also expected to be adapted. The key features are divided in two groups: functional abilities and system properties.

5.3.2.1 Functional abilities

*Functional abilities* include specific functionality that must be implemented on the robot in order to perform decently in the tests. Each test requires a certain subset of these abilities, as they are also directly represented in the score system. Teams must thus decide which of these abilities to implement and what degree of performance to achieve, depending on their background and the kind of tests they intend to participate in. *Functional abilities* currently are:

- *Navigation*, the task of path planning and safely navigating to a specific target position in the environment, avoiding (dynamic) obstacles
- *Mapping*, the task of autonomously building a representation of a partially known or unknown environment on-line

Figure 5.2: The intelligent machines that participated in 2009 in Graz, Austria.
5.3. The @HOME initiative

- **Person recognition**, the task of detecting and recognizing a person
- **Person tracking**, the task of tracking the position of a person over time
- **Object recognition**, the task detecting and recognizing (known or unknown) objects in the environment
- **Object manipulation**, the task of grasping or moving an object
- **Speech recognition**, the task of recognizing and interpreting spoken user commands (speaker dependent and speaker independent)
- **Gesture recognition**, the task of recognizing and interpreting human gestures.

5.3.2.2 System properties

System properties include demands on the entire robotic system that we consider of general importance for any domestic service robot. They can be described as “soft skills” which must be implemented for effective system integration and successful participation in the @HOME competition. Initial system abilities are:

- **Ease of use** - Laymen should be able to operate the system intuitively and within little amount of time.
- **Fast calibration and setup** - Simple and efficient setup and calibration procedures for the system.
- **Natural and multimodal interaction** - Using natural modes of communication and interaction such as, e.g. using natural language, gestures or intuitive input devices like touch screens.
- **Appeal and ergonomics** - General appearance, quality of movement, speech, articulation or HRI.
- **Adaptivity / General intelligence** - Dealing with uncertainty, problem solving, online learning, planning, reasoning.
- **Robustness** - System stability and fault tolerance.
- **General applicability** - Solving a multitude of different realistic tasks.
Although some of these properties cannot be benchmarked as directly as the functional abilities, they are considered as integral and an implicit part of the competition and tests.

The sorting mechanism (the ROBOCUP@HOME competition) is defined in such a manner that it allows for the generation of many types of solutions, instead of restricting them. The definition allows that the generators (the human teams and their robotic creations) remain in flux. Although this process consumes a lot of energy, it should lead to the next singularity.

5.3.3 Implementation of benchmarks

In the following, we are going to elaborate on how we implemented the ROBOCUP@HOME competition as a set of benchmark tests for service robots in domestic environments. This implementation is based on the concepts mentioned in the previous section.

The competition is organized in a multi-stage system. All qualified teams (currently up to 24) participate in the first stage called Stage I. It consists of a set of benchmarks with a focus on testing basic tasks and checking for a small set of key features with a limited amount of uncertainty involved. Then, the ten best teams advance to the second stage called Stage II where the benchmarks are more demanding, more realistic and involve more uncertainty and a higher level of system integration. In the Finals, the performance of the five best teams is evaluated by a jury. A combination of the jury score and the previous score from Stage I and Stage II determines the ranking.

The tests themselves comprise realistic and useful tasks for a domestic service robot. Each test evaluates certain key features. A tabular overview of the functional abilities required in each test can be found in Table 5.1. An overview of benchmarks where certain system properties are tested is given in Section 5.4.1.2. The implementation of the competition described in the following reflects the situation of the competition in 2008.

5.3.3.1 Score system

Two types of tests exist: Regular tests are specified in terms of the task to solve and the scoring. In open tests, teams can either freely choose what to show (the Open Challenge
and the Finals), or a topic is given according to which teams can do a demonstration (Demo Challenge). Since the scoring in the open tests is based on an evaluation by a jury, it is partially subjective. However, for every open test there is a list of criteria, which the jury bases its decision on. The criteria will be discussed in the test descriptions below. The scoring in the regular tests mainly reflects the key features mentioned earlier.

It is important not only to steer the generative mechanisms, but to also to allow for the generation of solutions that might show alternative paths through the phase space. Since it is uncertain which paths will lead to solutions that will create viable outcomes in the future, this freedom is essential. It leads to many solutions which are not viable at all, but once a fit (partial) solution to a (partial) problem has been found many generators (teams) copy this. This convergence towards local optima is counteracted by changes in the environment (tests and locations in the competition).

To keep the entry level for the competition reasonably low, while still aiming for high top level performance, a so-called partial score system was introduced in 2008. With this, a team receives a part of the total score for showing a part of the task’s specification. Each of the partial scores is connected to one or more of the functional abilities and/or system properties. This does not only allow for assessing the fulfillment of these features individually, but it is also an incentive for teams to participate in a test even if they know that they cannot solve it completely.

For each test in Stage I a maximum of 1000 points, and in Stage II 2000 points, can be scored.

5.3.3.2 Stage I Tests

The overall theme of tests in Stage I is to benchmark essential abilities and properties that any robot in @HOME should exhibit. During the setup days before the competition, a set of ten randomly chosen and previously unknown objects is provided to the teams. A subset of these objects is then used for certain tests.

Introduce In the Introduce test, a robot has to autonomously enter the scenario and move to a position in front of the audience. There, the robot has to introduce itself and the team using speech, gestures, slides or multimedia. Afterwards, it must leave the scenario on its own.
**Fast Follow**  A robot’s task in the Fast Follow test is to follow a person from one entrance of the environment to the other. Two teams compete against each other in this test starting from opposite ends. They need to pass by one another on a common path through the scenario. The most important capabilities evaluated in this test are detection and tracking of humans and safe navigation in a dynamic environment, a task which naturally includes obstacle avoidance.

**Fetch & Carry**  In the Fetch & Carry test, the robot has to find and retrieve a certain object which it then needs to return to the human instructor. The robot is instructed to get the object using natural language. Teams are allowed to give the robot a hint on the item’s location. Thus, speech recognition and natural language processing are essential to succeed in this test. Human-robot interaction by means of joint activities in common physical space is emphasized, since for the robot to understand the hint, it needs to be capable of interpreting the given spatial description.

The use of natural language is useful in two distinctive manners. The first is the ease of guidance and interaction of the developing AI. The second is that it forces the AI to adjust to humans, instead of humans to the machine.

**Who is Who?**  The main theme in the Who is Who? test is the detection and recognition of people. The robot has to find three persons (two of them unknown to the robot) spread out around an area near to the entrance of the scenario. The robot then needs to find these people, introducing itself to every person found. Each person has to be either identified (if known already) or learned (if unknown). Figure 5.3 on the left shows an example from the 2007 competition.

**Competitive Lost & Found**  For the Competitive Lost & Found test, two teams compete against each other directly. The assignment is to find and identify as many out of three objects as possible. The referees pick the objects randomly from the set of objects and distribute them randomly throughout the scenario (in a way that the robot actually has a chance of finding the object) just before the test starts.

**Open Challenge**  In order to allow all participating teams to freely demonstrate their scientific achievements and their unique robot features or capabilities the Open Chal-
5.3. The @HOME initiative

Figure 5.3: Team UT Austin doing Who is Who (left), team eR@sers teaching in an object (right)

Stage I concludes. Here, no restrictions on the kind of performance, kind of devices, or kind of interaction are applied. The Open Challenge is meant as a means of iteratively enhancing the @Home competition by integrating relevant and innovative aspects demonstrated by the teams in future tests. This test consists of a presentation given by the team (an example is depicted in Figure 5.4) and a demonstration of their robot. The performance is then evaluated by all other teams according to a list of predefined criteria. These criteria are as follows:

**Presentation** The quality of the presentation is evaluated as an indication of attractiveness, and to foster scientific exchange between the teams.

**Social relevance / Usefulness for daily life** Because RoboCup@Home is about socially relevant robotic applications, this aspect is also evaluated.

**Human-Robot interaction** The quality of human-robot interaction in the demonstration is evaluated according to the judging teams’ own focus and interests. This cross-evaluation should thus reflect the diverse community and give broad feedback for the performing team.

**Autonomy** The amount of autonomy shown by the robot during the demonstration. This is to avoid open loop behavior.

**Difficulty and success** The level of difficulty and success of the robot performance is evaluated.
Appeal/Relevance for @Home How well does the demonstration fit the scope of the @Home initiative? Could elements of the demonstration be integrated in future @HOME competitions?

Scientific value / Jury questions What is the scientific value of the approach, and how well did the team respond to jury questions?

The open challenge, just as the finals, is set up in such a manner that the sorting machine is less strict than with the other tests. Less restriction allows to explore generated solutions which might not have been thought of by the creators of the competition. The open challenge is a feedback mechanism on the performance of the competition, sorting out poorly functioning aspects of the sorting mechanism itself. It is also the most interesting part of the competition to watch and study, because of the relaxed constraints. The open challenge and the finals are used as a part of the abstract searching mechanism to generate new and possibly higher level sorting machines. The new sorting machines are used during the next iteration of the iterative search process.

Figure 5.4: Presentations from team Pumas(left) and team PAL(right) during the Open Challenge
5.3. The @HOME initiative

5.3.3.3 Stage II Tests

In contrast to the challenges in Stage I, Stage II comprises tests that are more complex, involve more uncertainty, and which check for the integration of several features in a more realistic, application-like setting.

*PartyBot*  The *PartyBot* test is an elaborated version of the *Who is Who?* test from Stage I where the robot’s task is to find, recognize, and/or remember multiple unknown people randomly distributed throughout the entire environment (standing and sitting) and to tell them apart later on when serving a drink.

*Supermarket*  To mimic the possible future application of assisted shopping the *Supermarket* test was introduced. The robot needs to retrieve certain household objects from a shelf for a person randomly chosen from the audience that does not know how to operate the robot. In 2010 this test is performed in a real supermarket. This demands the robot has the ability to explain its own modes of operation and to report on
the robot’s internal models to a layman. Furthermore, it requires speaker-independent speech recognition, and it enforces the ease of use proclaimed as a system property, since the operation by laymen raises uncertainty both in input and reaction.

**Walk & Talk** The task of the *Walk & Talk* test is to introduce a robot to a new environment and make it remember a set of places. A human leader guides the robot through the scenario that was completely rearranged beforehand (and is therefore unknown to the robot) and has to teach specific locations only using natural language. The robot then has to prove that it has correctly learned those locations by having to navigate to certain places in random order after a speech command is given.

**Cleaning Up** In the *Cleaning Up* test, a robot needs to collect a set of five unknown objects (i.e., not from the set of known objects) dispersed throughout the scenario. Objects can be anything that can be expected to lie around in a household. Restrictions are put on the size so that objects are not too small to be overlooked and not too big so they can still be handled by a robot by pushing or grasping.

**Demo Challenge** The *Demo Challenge* is an open demonstration similar to the *Open Challenge*, as no restrictions on the kind of interaction or the kind of external devices are applied. In contrast to the *Open Challenge* the topic of the demonstration is pre-defined and varies from year to year. It is meant to foster development in a certain area or on a particular theme with a strong relation to real applications and daily-life situations. It should provide a showcase of the current state of the art in home robotics and inspire both the community and the public. In 2008 the theme was “cooking”, i.e., the robot should assist a human in preparing a meal. The task was not formulated in any concrete specification. Possible means to assist were, for example, fetching a recipe from the Internet and retrieving ingredients necessary for the same. Figure 5.6 (left) shows a robot participating in the 2008 demo challenge.

The demo challenge functions as an exploration device with a little more focus than the open challenge and the finals. The competition thus has steering elements, feedback mechanisms, generative capabilities and is in (semi-)constant flux, changing parts of the tests on a yearly basis. This is akin to the circadian rhythm of the machine learning of the previous chapter.
5.3. The @HOME initiative

Figure 5.6: HomeBreakers robot Bender assisting cooking (left) and showing emotions (right)

5.3.3.4 Finals

The competition concludes with the Finals, where as in the Open Challenge, each team can demonstrate what they think is an important feature or capability of their robot. The idea, however, is to present a coherent story-like performance which is evaluated by an external jury according to a list of predefined criteria. Because teams that have reached the Finals have already proven to fulfill a variety of abilities, the criteria of the evaluation are slightly different from those in the Open Challenge.

**Scientific contribution / Contribution to the community** Amount, relevance and quality of the team’s contribution to the @Home community

**Relevance for ROBOCUP@HOME/ Usefulness for daily life** of the demonstration

**Usability / Human-robot interaction and multimodality** Ease of use, quality of HRI and multimodality during the demonstration

**Originality and presentation** Originality of the demonstration, quality of the presentation
Difficulty and success of the demonstration

Previous performance during Stage I and Stage II (determined by previous score)

The competition is experimenting with the sorting machine. This co-evolutionary process is essential to avoid locally optimized solutions. The finals are an implementation of the abstract searching mechanism actively exploring the phase space of paths towards an intelligent domestic service robot.

5.4 Evaluation and discussion

Two important objectives for an annual scientific competition are to provide a common benchmark to many teams, which allows for the measurement of performance advances over time, and to develop relevant scientific solutions and results. In this section we describe and discuss the results obtained by the ROBOCUP@HOME teams both in terms of performance in the tests and in terms of scientific achievements.

As for a team’s performance, it is important to note that the score system of ROBOCUP@HOME relates the desired abilities of the robots with the scores of the competition. In contrast to other competitions (e.g., RoboCup soccer), where the score hides many factors, the @HOME score provides an actual way of measuring the performance of teams in terms of such abilities. This score consequently enables an analysis of performance in order to update the rules and drive technological and scientific progress.

In the remainder of this section, first, we will present an analysis of 2008 team performance based on the relationship between key features and test scores; second, we will discuss the evolution of the league over time; then, we will highlight the teams’ main scientific contributions related to @HOME; and finally, we will discuss results from the @HOME community.

5.4.1 Representation of key features in the benchmarks

In the following the representation of key features, i.e. the functional abilities as well as the system properties, in the benchmarks and in the competition score are shown.
5.4. Evaluation and discussion

5.4.1.1 Functional abilities

Table 5.1 relates the functional abilities defined in Section 5.3.2 with the tests described above. It quantifies the maximum score distribution per test with respect to the contained functional abilities. For ease of notation, The following abbreviations are used. Tests include Fast Follow (FF), Fetch & Carry (FC), Who is Who (WW), Lost & Found (LF), PartyBot (PB), Supermarket (SM), Walk & Talk (WT), and Cleaning Up (CL). The abilities are Navigation (Nav), Mapping (Map), Person Recognition (PRec), Person Tracking (PTrk), Object Recognition (ORec), Object Manipulation (OMan), Speech Recognition (SRec), and Gesture Recognition (GRec). Note that for the Introduce test, the Open Challenge, the Demo Challenge, and the Finals, values are not indicated because teams can freely choose their performance and the focus on certain abilities themselves. This way, we expect new abilities to be demonstrated, which can be used to enhance the competition in the future.

Since the competition involves mobile robots, navigation is currently the most dominant ability represented in the score. Object manipulation and recognition also play an important role since service robots are useful if they can effectively manipulate objects in the environment. Person recognition, tracking, and speech/gesture recognition are needed to implement effective human-robot interaction behaviors. As gesture recognition was introduced as a new (and optional) ability in 2008, its weight in the total score is still comparably low. Finally, mapping plays a more limited role; such an ability is used in the Walk & Talk test, where the environment is completely remodeled during the test, so the robot enters in an unknown environment, while for other tests only minor modifications of the environment are made right before the tests. Thus, pre-computed maps (either built off-line by the robot or manually drawn) can be used.

This table is important in order to define the weight of each ability in a test and in order to distribute the abilities among the tests. Furthermore, one can analyze the performance of the teams and the difficulty of the tests after a competition. This allows for an iterative and constant development of the benchmarks.
Chapter 5. Gaseous interactions

Table 5.1: Distribution of test scores related to functional abilities

<table>
<thead>
<tr>
<th>Test</th>
<th>Nav</th>
<th>Map</th>
<th>PRec</th>
<th>PTrk</th>
<th>ORec</th>
<th>OMan</th>
<th>SRec</th>
<th>GRec</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>550</td>
<td>0</td>
<td>0</td>
<td>450</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>FC</td>
<td>375</td>
<td>0</td>
<td>0</td>
<td>150</td>
<td>400</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>WW</td>
<td>350</td>
<td>0</td>
<td>550</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>LF</td>
<td>550</td>
<td>0</td>
<td>0</td>
<td>450</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>PB</td>
<td>1000</td>
<td>0</td>
<td>700</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>0</td>
<td>2000</td>
</tr>
<tr>
<td>SM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>400</td>
<td>1000</td>
<td>200</td>
<td>400</td>
<td>2000</td>
</tr>
<tr>
<td>WT</td>
<td>918</td>
<td>416</td>
<td>0</td>
<td>250</td>
<td>0</td>
<td>0</td>
<td>416</td>
<td>0</td>
<td>2000</td>
</tr>
<tr>
<td>CL</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>550</td>
<td>450</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2000</td>
</tr>
<tr>
<td>Tot</td>
<td>4743</td>
<td>416</td>
<td>1250</td>
<td>700</td>
<td>1550</td>
<td>2150</td>
<td>791</td>
<td>400</td>
<td>16000</td>
</tr>
</tbody>
</table>

The progress of the competition is improved by deliberatively focusing on difficult topics. Higher levels of AI are reached by incorporating several sorting machines into a single new sorting machine. An example of this chunking of sorting machines (building a meshwork of sorting machines which operate as a single new one) is the combination of navigation, manipulation and person recognition capabilities in a test where the robot has to serve a drink to a person.

5.4.1.2 System properties

Similar relationships between system properties and the tests exist. As previously mentioned, this relationship cannot be quantified in scores as easily, as the system properties are of more implicit meaning for the tests. However, on the basis of the objective of the tests, the importance of each of the system properties can be estimated. Table 5.2 relates tests with system properties by denoting a 'very important' relation with '+'+, an important relation with '+', and a minor relation with '-'. Note that these symbols are used only to indicate the importance of system properties in a test, rather than defining the score of a test.

System properties are further represented in the general rules, in overall requirements, and in special properties in certain tests. By using laymen to operate the robots in the Supermarket test, the Who is Who test, and the PartyBot test, Ease of Use (EUse) is enforced. The restrictions on setup time and procedures demands for Fast Calibration and Setup (FCal). Natural Interaction (NInt) and Multimodal input is rewarded in
5.4. Evaluation and discussion

<table>
<thead>
<tr>
<th>Test</th>
<th>EUse</th>
<th>FCal</th>
<th>NInt</th>
<th>App</th>
<th>Adap</th>
<th>Rob</th>
<th>GAppl</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FF</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>WW</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>LF</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>OC</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>PB</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>SM</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>WT</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>CL</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Dem</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>++</td>
</tr>
<tr>
<td>Fin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>++</td>
</tr>
</tbody>
</table>

Table 5.2: Importance of system properties in each test

the Supermarket test.

*Appeal and Ergonomics* (App) are part of the evaluation criteria in the Introduce test, the Open Challenge, and the Finals. *Adaptivity* (Adap) is especially important in the Cleaning Up test. The limited number of specifications in the tests and the environment, and the fact that people who interact with the robot are chosen randomly in many tests, demands *Robustness* (Rob).

Finally, a team can only reach the *Finals* if its robot performs well in many tests with different tasks to solve. This incorporates the aspect of *General Applicability* (GAppl).

5.4.2 Analysis of 2008 team performance

In the following, we will analyze the performance of the teams in these abilities during *ROBOCUP@HOME 2008* competition.

Table 5.3 presents the scores actually gained by the teams during the competition and the percentage with respect to the total score available, related to each of the desired abilities. The third column shows the result obtained by the best team, while the fourth one is the average of the results of the five finalist teams. This table allows for many considerations, such as: 1) which abilities have been most successfully implemented by the teams? 2) how difficult are the tests with respect to such abilities? 3) which tests
Table 5.3: Available and achieved score for the desired abilities

<table>
<thead>
<tr>
<th>Ability</th>
<th>Available score</th>
<th>Achieved score [max]</th>
<th>Achieved score [avg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>4743 (40%)</td>
<td>1892 (40 %)</td>
<td>1178 (25 %)</td>
</tr>
<tr>
<td>Object Manipulation</td>
<td>2150 (18%)</td>
<td>75 (3 %)</td>
<td>15 (1 %)</td>
</tr>
<tr>
<td>Object Recognition</td>
<td>1550 (13%)</td>
<td>450 (29 %)</td>
<td>125 (8 %)</td>
</tr>
<tr>
<td>Person Recognition</td>
<td>1250 (10%)</td>
<td>400 (32 %)</td>
<td>190 (15 %)</td>
</tr>
<tr>
<td>Speech Recognition</td>
<td>791 (7%)</td>
<td>692 (87 %)</td>
<td>293 (37 %)</td>
</tr>
<tr>
<td>Person Tracking</td>
<td>700 (6%)</td>
<td>700 (100 %)</td>
<td>570 (81 %)</td>
</tr>
<tr>
<td>Mapping</td>
<td>416 (3%)</td>
<td>416 (100 %)</td>
<td>183 (44 %)</td>
</tr>
<tr>
<td>Gesture Recognition</td>
<td>400 (3%)</td>
<td>0 (0 %)</td>
<td>0 (0 %)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12000 (100%)</strong></td>
<td><strong>4909 (41 %)</strong></td>
<td><strong>2554 (21 %)</strong></td>
</tr>
</tbody>
</table>

and abilities need to be changed in order to guide development into desired directions?

Table 5.4 summarizes the number of teams participating in each test and those which received a non-zero score. This table helps to evaluate team preferences and difficulty of the tests. Note that teams were not required to perform all the tests. Therefore, some of the zero scores in the table derive from a team’s choice not to participate in a test.

An evaluation of system properties is more complicated since they are difficult to quantify precisely. How to do this appropriately is a research track in ROBOCUP-@HOME in itself.

One of the implications of the theory of GAI is that the search mechanism is worth of scientific study. The answer to the question of how to incrementally create successive searching machines with increased complexity that structure generative mechanisms is essential for the automated creation of intelligent machines.

Our current approach is to test for system properties through general requirements and to enforce the combination of functional abilities.

An analysis of these results is very helpful for the future development of the @HOME competition. It gives direct, quantitative feedback on the performance of the teams with respect to key abilities and tasks. This allows us to identify abilities and respective tests which need to be modified, and to adjust the weights of certain abilities with respect to the total score. Possible modifications involve:

- Increasing the difficulty if the average performance is already very high
5.4. Evaluation and discussion

<table>
<thead>
<tr>
<th>Test</th>
<th>Participating Teams</th>
<th>Teams with non-zero score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduce</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Fast Follow</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Fetch &amp; Carry</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Who’s Who</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Comp. Lost &amp; Found</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Open Challenge</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Party Bot</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Supermarket</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Walk &amp; Talk</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Robot Chef</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cleaning</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 5.4:** Number of teams participating and gaining score for each test.

- Merging abilities into high-level skills, more realistic tasks
- Maintaining or even decreasing difficulty if the observed performance is not satisfying
- Introducing new abilities and tests

As the integration of abilities will play an increasingly important role for future general purpose home robots, this aspect should especially be considered in future competitions.

The focus on the sorting mechanisms demonstrates the difficulty of the entire process of generating these sorting machines. From this perspective ROBOCUP@HOME acts as a generator of sorting machines.

5.4.3  **League progress**

The results obtained so far by the @HOME initiative can be measured on several levels:

- increased number of participating teams and of community members,
• increasing performance in the tests,
• increase of public awareness (media, press, Internet),
• increasing number and quality of scientific contributions (see next section).

For some of these measures, a quantitative analysis over the years is presented in the following.

Since 2006, a total of 25 teams distributed worldwide (12 from Asia, 8 from Europe, 4 from America, 1 from Australia), have participated in the three years of the ROBO-CUP@HOME world championship. Furthermore, national competitions have been established in China, Mexico, Germany, Iran and Japan. These events are useful not only to test team developments and rules, but also to possibly select teams that will participate to the world championship.

Table 5.5 describes the number of participating teams in the annual world championship. The second column shows the number of teams that pre-registered and delivered the necessary qualification material, such as videos and a team description paper. The third column shows the number of teams that qualified after a review from the Organizing Committee, and the fourth column shows the number of teams that actually participated in the competitions. Finally, the fifth column shows the number of new teams (i.e., teams that did not participate in the previous years). The last line refers to the 2009 competition, for which 21 teams from 14 countries participated. Figure 5.7 shows the participating @HOME teams at the RoboCup world championship 2008 in Suzhou, China.

<table>
<thead>
<tr>
<th>Year</th>
<th>Pre-registration</th>
<th>Qualification</th>
<th>Participation</th>
<th>New teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>12</td>
<td>12</td>
<td>12 (440; 2.72%)</td>
<td>12</td>
</tr>
<tr>
<td>2007</td>
<td>16</td>
<td>13</td>
<td>11 (321; 3.42%)</td>
<td>5</td>
</tr>
<tr>
<td>2008</td>
<td>18</td>
<td>17</td>
<td>14 (373; 3.75%)</td>
<td>8</td>
</tr>
<tr>
<td>2009</td>
<td>26</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.5: Number of participating teams

The number of and the increase in participating teams must be also related to the general participation across all leagues. (The number of total teams and percentage of @HOME teams are given in parenthesis in the fourth column). Regardless of the drop in the total number of teams throughout all leagues in 2007 (in the US) and 2008
(in China), mainly due to high travel and shipping costs, as well as the difficulties in custom and visa affairs, the increase of percentage of @HOME teams is a clear indication of the growth of the league.

Another important parameter to assess the results of the competition is the increase in performance. Obviously, it is difficult to determine such measure quantitatively. The main reason is that the constant evolution of the competition and the iterative modification of both the rules and the partial scores do not allow for a direct comparison.

However, it is possible to identify certain situations which indicate the success of the initiative in terms of general performance increase. Table 5.6 gives some examples for this increase over the last three years. The first row contains the percentage of unsuccessful tests, i.e., tests where no score was achieved at all, dropping from 83% in 2006 to 41% in 2008. The second row shows the increase in the total number of tests per competition. The third row indicates the average number of tests that teams participated in successfully (i.e., with a non-zero score). The enormous increase from from 1.0 tests in 2006 to 4.9 in 2008 is a strong indication of an average increase in robot abilities and in overall system integration.
<table>
<thead>
<tr>
<th>Measure</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of 0-score performance</td>
<td>83%</td>
<td>64%</td>
<td>41%</td>
</tr>
<tr>
<td>Total number of tests</td>
<td>66</td>
<td>76</td>
<td>86</td>
</tr>
<tr>
<td>Avg. number of succ. tests p. team</td>
<td>1.0</td>
<td>2.5</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Table 5.6: Measures indicating general increase of performance

The little table above (table 5.6) is perhaps the most important one in this chapter. Although the tests and scenario become more complex over the years, the increase in performance demonstrates that there is successive scaffolding of the structures (robots and software) by the generative mechanisms (the teams and their scientific work). According to GAI it is because of the large flux of the ROBOCUP@HOME league that robust structures are generated. Their robustness stems from the focus on general internal mechanisms of the machines in order to avoid local optima. This active avoidance of the local optima while aiming for a goal is probably the single most important aspect of GAI and of ROBOCUP@HOME. They are both goal directed, even when the goal is not well-defined. The goal of GAI is to the automated creation of intelligence, the goal of ROBOCUP@HOME is an intelligent domestic service robot. The same holds for natural beings. Their goal is to produce offspring that procreates, which is also not well-defined. Still, all these processes are goal oriented.

5.4.4 Scientific achievements

In addition to numerical analyses of test performances, relevant scientific achievements have been obtained by teams participating in the competition. ROBOCUP@HOME provides a proper setting for developing and testing integrated solutions for mobile service robots. As a result, robot hardware and software architectures evolve over time.

This effort is demonstrated in scientific papers and in the teams’ reports (Team Description Papers), which contain technical and scientific details on the hardware/software architectures and the implemented approaches and functionality. In particular, due to the nature of the @HOME competition, in these architectures special focuses are put on Human-Robot-Interaction (e.g. (Savage et al., 2008)), on personal assistive robots (e.g. (Ruiz-del-Solar, 2007)) and on high level programming for domestic ser-
vice robots (e.g. (Schiffer et al., 2006)).

Scientific advancements can be also identified in specific functionality. Speech recognition evolved from difficult interaction with headsets and portable laptops (2006-2007) to speaker-independent speech recognition with effective noise cancellation using onboard microphones (2008) (Doostdar et al., 2008). Face recognition has been made robust in the presence of spectators standing around the edges of the scenario (Correa et al., 2008; Knox et al., 2008) and tuned for real-time use (Belle et al., 2008) (Figure 5.3 left). Object recognition in @HOME requires a more general approach than the color-based recognition used in the soccer leagues, and it offers a challenging testbed. Techniques using different feature extractors and different matching procedures have been tested (e.g. (Loncomilla and Ruiz-del-Solar, 2007)), reaching a level in which the robot can reliably remember an object shown by a user (by holding it in front of the robot) and then recognize it among several others (2008, Figure 5.3 right)). Gesture detection and recognition has also been studied in order to communicate with the robot, and uses an effective approach based on active learning (Francke et al., 2007). Finally, object manipulation has evolved from gathering a newspaper from the floor (2006), to grasping cups from a table (2007), to grasping different objects at various heights (2008) (Figure 5.5).

A measure of the scientific contributions is also given by the five papers (out of 56) related to ROBOCUP@HOME presented to the International RoboCup Symposium 2008, including one that received the best student paper award (Doostdar et al., 2008). In comparison with all the RoboCup leagues and sub-leagues, @HOME ranked third out of ten with respect to the number of papers presented at the RoboCup Symposium (together with Soccer Middle-Size and Soccer Simulation).

5.5 Conclusion and Outlook

This chapter presented the ROBOCUP@HOME initiative as a community effort to develop and benchmark domestic service robots through scientific competitions. To do so, we employ so-called system benchmarking that evaluates a robot’s performance in a realistic, complex and dynamic environment. The general setting is designed to exhibit a high degree of uncertainty that the robots have to deal with.

The rules of the competition aim to implement the benchmark by means of general
rules and a set of specific tests. Evaluation is conducted along a set of key features. These features, divided into functional abilities and system properties need to be met in order to be successful in the competition. The modular and open character of the competition’s framework allows for an iterative adaptation of features and tests according to the observed and measured benchmark performances.

Special focus is put on establishing a community to foster interdisciplinary exchange of knowledge and technology. Furthermore, this community is essential to create common vision and understanding for the problems and goals of the @HOME initiative, and to give feedback for the iterative development of the competition.

Starting with the first competition in 2006, the overall development of the initiative with respect to performance increases, the growing community, knowledge exchanges and public awareness has been very promising over the past three years. @HOME has become the largest international competition for domestic service robots, with currently five national competitions in China, Japan, Germany, Iran and Mexico besides the annual world championships. Competitions in South America and the US are expected to be introduced in 2009.

The future development of the @HOME competition is highly iterative, as it involves constant feedback from the community, adjustments on the focus of desired abilities and changes of the rules.

In general, the tests, functional abilities and desired system properties will evolve over the years and will be combined to form more realistic high-level tasks. New tests with different focuses and higher complexity will be added in the future, depending on the results of previous years.

Still, short, mid and long-term goals are necessary, as they help identify and approach the problem in the large, real-world problem space in a structured way. At the moment the focus is on physical capabilities such as manipulation, human recognition and navigation. In the future, more focus will be put on artificial intelligence and mental capabilities in the context of HRI. This includes situational awareness, online learning, understanding and modeling the surrounding world, recognizing human emotions and having appropriate responses.

The increase of complexity in the competition from 2007 to 2008 was rather high. Therefore, the Technical Committee of the @HOME league agreed to make only minor modifications to the rules in 2009. In 2010 we are planning for the first tests in outside
of the scenario, for example a supermarket and a shopping mall. Rule changes for next year will involve an increased focus on HRI, e.g. combined use of speech and gestures, robot operation by laymen, or following previously unknown persons. Application scenarios will become more realistic, e.g. the demo challenge will involve robots serving drinks and food at a real party setting involving many people unfamiliar with the robots. Furthermore, uncertainty and dynamics in the environment are increased by changing object positions more frequently, having more people in the scenario, and leaving the scenario with the robots.

Further, an annual @HOME camp will be established in 2010. It will consist of a set of lectures and practical sessions from and for members of the community. Having a separate event exclusively for knowledge exchange in the absence of any competitive aspect will help to foster exchange of knowledge even more. Also, new research groups and communities will be addressed and invited to join and share their knowledge with the @HOME community.

Midterm goals include the search, identification, design and use of a common robot software architecture or framework to better exchange and reuse software components already developed in the community and beyond. The same holds true for hardware, where companies or groups with relevant hardware components like sensors, actuators, or even standard robot platforms will be identified and asked to join and support the community.

The future @HOME scenario will contain more ambient intelligence, with which the robots can interact. The use of the Internet as a general knowledge base, and the communication with household devices, TVs, or external video cameras are some examples.

Another midterm goal is gradually testing the robots in the real world, e.g. going shopping in a real supermarket or taking public transportation. Moreover, usability and appearance of the robots will be of higher importance if one wants to increase their public acceptance.

In general, the competition will move toward a high-level integration of the identified abilities into more realistic and relevant applications. This will increase attractiveness, generate more public awareness and hopefully inspire and accelerate consumer product development for domestic service robotic applications in the near future.