Using Heuristic Evaluation for Human-Humanoid Robot Interaction in the Soccer Robotics Domain

Mohan Rajesh Elara[#1], Carlos A. Acosta Calderon[#2], Changjiu Zhou[#3], Pik Kong Yue[#4], Lingyun Hu[#5]

[#] *Advanced Robotics and Intelligent Control Centre,*
School of Electrical and Electronics Engineering,
Singapore Polytechnic, 500 Dover Road, Singapore
URL: www.robo-erectus.org
{ZhouCJ, MohanRajesh}@sp.edu.sg

Abstract—**In the field of human-computer interaction (HCI), one of the most popular inspection-based methods for evaluating usability is the Heuristic Evaluation (HE) in gauging and improving the interaction design. With mobile robots exhibiting a stronger presence in commercial markets in the last few years, very little work has been done with heuristic evaluation in human-robot interaction discipline. Heuristic evaluation requires application of an established set of heuristics (guidelines) as they review a given system. This paper focuses on deriving a set of heuristics for use with human-humanoid robot interaction (HHRI) system in soccer robotics domain. Our derived set of HHRI heuristics can be readily adopted by humanoid soccer robotic researchers with little or no evaluation experience. Experiments were performed with the derived set of HHRI heuristics and adequate results were obtained exhibiting improvement in soccer performance of the humanoid robot.**

I. Introduction

Growing popularity and increasing viable application domains has contributed to greater presence of robots in the commercial marketplace. Among various robotic platforms, especially humanoid robotic research has seen a rapid growth in the recent years due to the ability of humanoid robots to behave and interact like humans. Humanoid robots might provide day-to-day support in the home and the workplace, doing laundry or dishes, assisting in the care of the elderly, or acting as a caretaker for individuals within a home or institution [1]. Many of these tasks will involve a close interaction between the robot and the people it serves. Researchers have studied the humanoid robot development, including control, emotional expressiveness, humanoid-humanoid collaboration, human-humanoid robot interaction (HHRI) and perception [2][3][4]. However, there have been only few studies on evaluation of human-humanoid robot interaction (HHRI).

Heuristic evaluation is one of the popular formative evaluation techniques extensively used and categorized as discount evaluation due to its lower requirements on time and cost by the human-computer interaction (HCI) community [5]. Heuristic evaluation is a usability evaluation method to identify usability problems in the user interface (UI) design. It specifically involves evaluators examining the interface and judging its compliance with recognized usability principles called heuristics. Therefore, use of heuristic evaluation to a particular application domain requires a set of customized heuristics to be derived for that domain.

This paper focuses on deriving a set of customized heuristics for use with human-humanoid robot interaction (HHRI) system in soccer robotics domain. Our derived heuristics are obtained by tailoring the Nielsen's original set heuristics [6] to suit the human-humanoid robot interaction (HHRI) domain. We applied the customized heuristics to the evaluation of our human-humanoid robot interaction (HHRI) system of Robo-Erectus Junior, a soccer playing humanoid robot for improved soccer performance. Results of our experiments showed that our customized heuristics performed better against Nielsen's original set of heuristics for human-humanoid robot interaction (HHRI) in soccer domain. Also, our customized heuristics passed the canonical test with 3-5 evaluators finding 40-60% of the known usability problems. Our derived set of HHRI heuristics can be readily adopted by researchers with little or no evaluation experience.

II. Heuristic Evaluation

Heuristic evaluation is a form of usability inspection where usability specialists judge whether each element of a user interface follows a list of established usability heuristics. Usually analysts evaluate the system with reference to established guidelines or principles, noting down their observations and often ranking them in order of severity. The analysts are usually experts in human factors or HCI, but others, less experienced have also been shown to report valid problems [7]. A heuristic or expert evaluation can be conducted at various stages of the development lifecycle, although it is preferable to have already performed some form of context analysis to help the experts focus on the circumstances of actual or intended product usage. Heuristic evaluation has four stages:

- The pre-evaluation period, in which specific heuristics that can be applied to interaction system under study are obtained by tailoring the general heuristics. A reference guide is prepared for the evaluators detailing system under study, and procedures for heuristic evaluation.
- The briefing session, in which the invited panel of evaluators are told individually on what to do during the evaluation.
- The evaluation period, in which each expert typically spends 1-2 hours independently inspecting the system,

using the heuristics for guidelines. The first round of evaluation gives the evaluators an overall idea on the interaction design. The second round of evaluation allows the evaluators to focus on specific interface elements and to identify potential usability problems.

- The debriefing session, in which the experts come together to discuss their findings and to prioritize the problems they found and suggest solutions. The participants in the debriefing should include the evaluators, any observer used during the evaluation sessions, and representatives of the design team. The debriefing session would be conducted primarily in a brainstorming mode and would focus on discussions of possible redesigns to address the major usability problems and general problematic aspects of the design. A debriefing is also a good opportunity for discussing the positive aspects of the design, since heuristic evaluation does not otherwise address this important issue.

Heuristic evaluation is performed by having each individual evaluator inspect the interface alone. Only after all evaluations have been completed are the evaluators allowed to communicate and have their findings aggregated. This procedure is important in order to ensure independent and unbiased evaluations from each evaluator. The output from using the heuristic evaluation method is a list of usability problems in the interface with references to those usability principles that were violated by the design in each case in the opinion of the evaluator. It is not sufficient for evaluators to simply say that they do not like something; they should explain why they do not like it with reference to the heuristics or to other usability results. The evaluators should try to be as specific as possible and should list each usability problem separately [8].

This method provides quick and relatively cheap feedback to designers. Result of the evaluation generates good ideas for improving the user interface. The development team will also receive a good estimate of how much the user interface can be improved. There is a general acceptance that the design feedback provided by the method is valid and useful. It can also be obtained early on in the design process, whilst checking conformity to established guidelines helps to promote compatibility with similar systems [9].

The original list of heuristics as defined by Nielsen is: (1) Visibility of system status, (2) Match between system and the real world, (3) User control and freedom, (4) Consistency and standards, (5) Error prevention, (6) Recognition rather than recall, (7) Flexibility and efficiency of use, (8) Aesthetic and minimalist design, (9) Help users recognize, diagnose and recover from errors, (10) Help and documentation. However, this list of heuristics cannot be directly applied for human-humanoid robot interaction (HHRI) system due to its control architecture, autonomous nature and dynamically changing environment.

A. Heuristic Evaluation for HHRI

In this section, we derive a set of heuristics based on Nielsen's original set of heuristics, but modified to be more applicable to humanoid soccer robotics domain. We eliminated the irrelevant heuristics from the original Nielsen's list of heuristics and added new heuristics thru independent review by a panel of four researchers working in interaction design and humanoid soccer robotics areas in our lab based on the objectives for humanoid robots. Table I shows the modified list of heuristics for human-humanoid robot interaction (HHRI) system.

TABLE I
MODIFIED HEURISTICS FOR HHRI SYSTEM

No.	Modified Heuristics
1.	**Visibility of system status** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2.	**Clarity in information presentation** The interface should be designed to present clear and understandable information on sensors and actuators. Interface design must emphasis on recognition rather than recall with 'just enough' information to determine if intervention is needed.
3.	**Match between system and the real world** The language of the interaction between the user and the system should be in terms of words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
4.	**Extendibility of the system** The system must support evolution allowing inclusion of additional sensors, actuators, behavioural and skill components.
5.	**Help users recognize, diagnose, and recover from errors** Any system errors must be expressed in plain language (no codes), precisely indicating the problem, and constructively suggesting a solution.
6.	**Effective communication architecture** The system must allow effective and efficient user to system, system to system and system to user communications.
7.	**Aesthetic and minimalist design** The system should not contain information that is irrelevant or rarely needed. The physical embodiment of the system should be pleasing in its intended setting.

The list of modified heuristics was further refined by rounds of surveys and group discussion with 15 research staff and students working on humanoid soccer robotics projects. Each survey participant was to provide a relevance rating on a scale of 1 to 5 (5 being highest). The participants were also encouraged to comment on the given set of heuristics. The average relevance ratings for each heuristics were greater than 4.35. The comments obtained from the survey and group discussion served as a source of improvement for the modified set of heuristics. For example, one survey participant commented that prioritized placement of control buttons on the interface based on functional importance and frequency of use was necessary. Another survey respondent commented that abstract information on the humanoid robot and its environment must be provided to the user for debugging purposes in case of an error. Table II shows the refined final list of heuristics for human-humanoid robot interaction (HHRI) system.

TABLE II
FINAL CUSTOMIZED HEURISTICS FOR HHRI SYSTEM

No.	Final Customized Heuristics
1	**Visibility of system status** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time. The system should represent its humanoid robot and world models to the user so that the user has a full knowledge of the status of humanoid robot with respect to the world.
2	**Clarity in information presentation** The interface should be designed to present clear and understandable information on sensors and actuators. Interface design must emphasis on recognition rather than recall with 'just enough' information to determine if intervention is needed.
3	**Match between system and the real world** The language of the interaction between the user and the system should be in terms of words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
4	**Prioritized placement of information** Prioritized placement of control buttons on the interface based on functional importance and frequency of use.
5	**Extendibility of the system** The system must support evolution allowing inclusion of additional sensors, actuators, behavioural and skill components.
6	**Help users recognize, diagnose, and recover from errors** Any system errors must be expressed in plain language (no codes), precisely indicating the problem, and constructively suggesting a solution. Abstract information on the humanoid robot its environment must be provided to the user for debugging purposes.
7	**Effective communication architecture** The system must allow effective and efficient user to system, system to system and system to user communications. In case of multi interface systems, switching between systems must be achieved with ease.
8	**Aesthetic and minimalist design** The system should not contain information that is irrelevant or rarely needed. The physical embodiment of the system should be pleasing in its intended setting.

III. Robo-Erectus Junior – A Humanoid

This section describes the Robo-Erectus Junior humanoid robot and its human-humanoid interaction systems that we evaluated with heuristic evaluation. The Robo-Erectus project is developed in the Advanced Robotics and Intelligent Control Centre of Singapore Polytechnic. Robo-Erectus Junior is one of the foremost leading soccer playing humanoid robots in the RoboCup Humanoid League. Robo-Erectus Junior won the 2nd place in the Humanoid Walk competition at the RoboCup 2002 and got 1st place in the Humanoid Free Performance competition at the RoboCup 2003. In 2004, Robo-Erectus Junior won the 2nd place in Humanoid Walk, Penalty Kick, and Free Performance. The aim of the Robo-Erectus Junior development team is to develop a low-cost humanoid platform. The development of Robo-Erectus Junior has gone through many stages in the design of its mechanical structure, electronic control system and gait movement control. Figure 1 shows the physical design of Robo-Erectus Junior. Robo-Erectus Junior has been designed to cope with the complexity of a 2 versus 2 soccer game. It has three processors each for vision, artificial intelligence and control. Table III shows the specification of the processors used in Robo-Erectus Junior. The platform is equipped with three sensors: an USB camera to capture images, a tilt sensor to detect a fall, and a compass to detect their direction [10]. The servomotors used send back the feedback data including angular positions, speed, voltage, and temperature. To communicate with its teammates, Robo-Erectus Junior uses a wireless network connected to the artificial intelligence processor.

The vision processor performs recognition and tracking of objects of interest including ball, goal, field lines, goal post teammate and the opponents based on a blob finder based algorithm. The further processing of detected blobs, wireless communications and decision making are performed by the artificial intelligence processor which selects and implements the soccer skills (i.e. walk to the ball, pass ball, kick ball, dive….) the robot is to perform. Finally, the control processor handles the low level control of motor based on the soccer skill selected by the artificial intelligence processor. Robo-Erectus Junior was fabricated to participate in RoboCup 2007 in the KidSize category. Table IV shows the physical specifications of Robo-Erectus Junior.

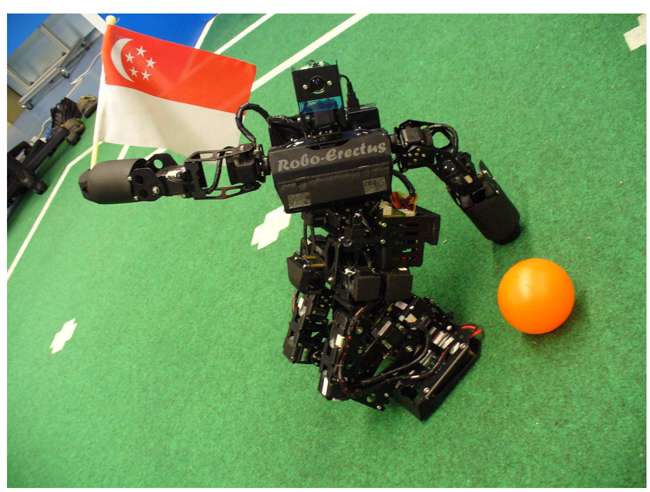

Fig. 1 Robo-Erectus Junior, the Latest Generation of the Family Robo-Erectus

TABLE III

PROCESSOR SPECIFICATION OF ROBO-ERECTUS JUNIOR

Features	Artificial Intelligence	Vision Processor	Control Processor
Processor	Intel ARM XScale	Intel ARM XScale	ATMEL ATmega-128
Speed	400Mhz	400Mhz	16Mhz
Memory	16MB	32MB	4KB
Storage	16MB	16MB	132KB
Interface	RS232, WIFI	RS232, USB	RS232, RS485

TABLE IV

PHYSICAL SPECIFICATION OF ROBO-ERECTUS JUNIOR

Weight	Dimension			Speed
	Height	Width	Depth	Walking
3.2 Kg	480 mm	270 mm	150 mm	2 m/min

It is powered by two high-current Lithiumpolymer rechargeable batteries, which are located in each foot. Each battery cell has a weight of only 110g providing 12v which means about 15 minutes of operation [11]. In the RoboCup 2007 competitions, Robo-Erectus Junior participated in the 2 versus 2 Soccer Games and the Technical Challenges. It was placed at 6th rank in the RoboCup 2007 humanoid 2 Vs 2 soccer games, Atlanta USA.

B. Human-Humanoid Interaction (HHRI) System

Figure 2 shows the human-humanoid interaction system of Robo-Erectus Junior, containing wealth of information. A separate window shows the video streaming from the robot. On the upper left of the interface is the score board showing the current status of the soccer match below which is the simulated actual match with the field localized Robo-Erectus Junior and its team mates. On the right is the control buttons for team configuration and game status signals.

Team configuration includes controls for assigning robot identification numbers, goal colours, and individual roles of robots whereas the game status signal includes information for the robot on the status of the match including kick off, time out, time off, resume, ball out, indirect kick and catch.

The purpose of performing heuristic evaluation for our human-humanoid interaction (HHRI) system is to uncover usability problems and thereby improve our humanoid robot performance in the soccer domain.

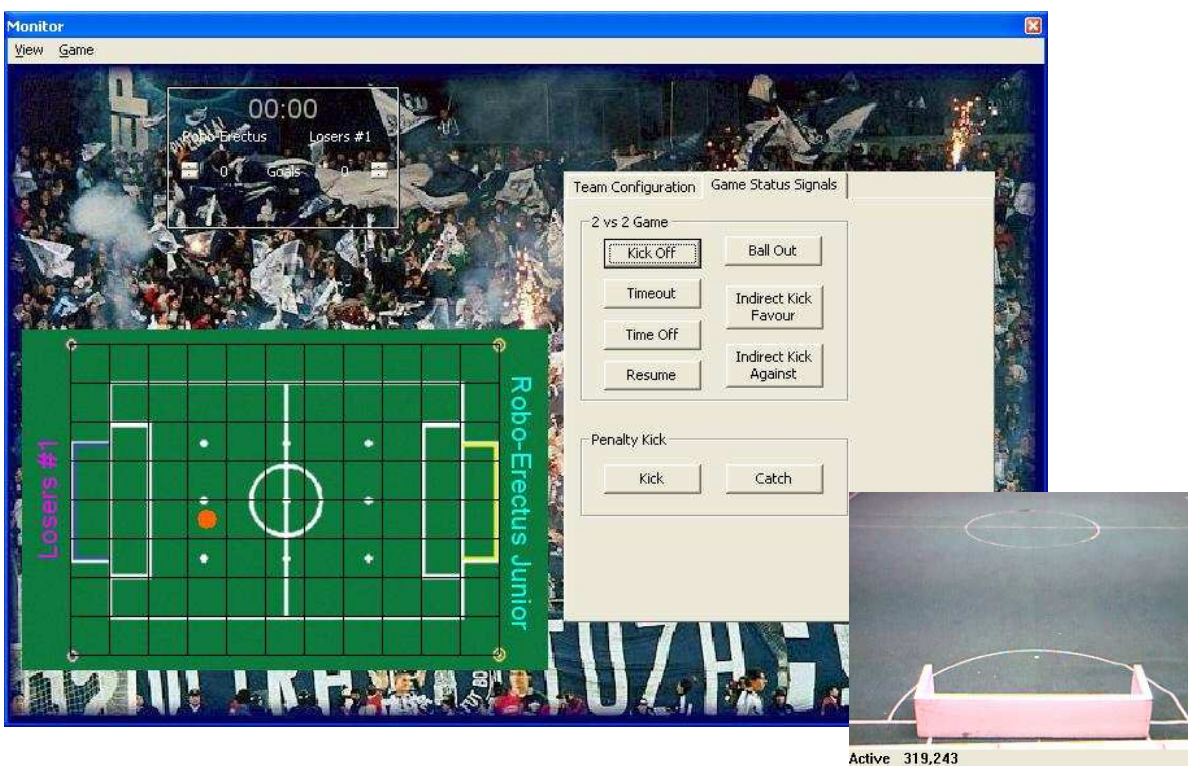

Fig. 2 Human-Humanoid Robot Interaction (HHRI) System of Robo-Erectus Junior Humanoid Robot

IV. Experimental Results

We recruited 9 research staff and students from our lab to evaluate the human-humanoid interaction (HHRI) system of Robo-Erectus Junior humanoid robot. The evaluators had a mean age of 22 years and two of them were female. All the evaluators are currently working on humanoid soccer robotics and human-robot interaction related areas. A planning session was conducted a month ahead of the experiment period to analyse the resources needed and issues to be addressed. During the briefing session for evaluators, information sheets providing details of the evaluation including, introduction to heuristic evaluation, humanoid platform, Robocup humanoid soccer league rules and regulations were distributed to the evaluators. All evaluators were given sufficient time to become familiar with the human-humanoid robot interaction system of Robo-Erectus Junior along with intended task scenarios. Each evaluator was given an evaluation period of one week for performing the heuristic evaluation with both Nielsen's original set of heuristics and final customized heuristics. The evaluators then worked with the system using mock tasks and recorded their observations as a list of usability problems. After the evaluation period, the evaluators collated the usability problem lists and rated the individual items for severity on a 5 point scale with 5 being the most severe usability problem and 1 being the least severe usability problem. Another independent evaluation was performed with expert evaluators to generate a master list of usability problems and their severities with Nielsen's original set of heuristics. A total of 36 usability problems are known for the human-humanoid interaction (HHRI) system for Robo-Erectus Junior and 32 of which were found in the heuristic evaluation using final customized heuristics with very low false positives. Some of the usability problems found include no proper display of robot-robot communication, inconvenient placement of control buttons, inflexible camera control and so on. Table V shows the number of known usability problems at each severity rating and compares the percentage of issues found with Nielsen's original set of heuristics and with the final customized set of heuristics.

TABLE V
USABILITY PROBLEMS ITS SEVERITIES AND PERCENTAGE OF PROBLEMS FOUND WITH NIELSEN'S SET OF HEURISTICS AND CUSTOMIZED HEURISTICS

Severity	Known Problems	% Problems Found with Nielsen's Heuristics	% Problems Found with Final Customized Heuristics
1	6	66.67%	66.67%
2	5	60%	80%
3	12	58.33%	91.67%
4	10	50%	100%
5	3	33.33%	100%

From the Table V, it is clearly evident that the percentage of usability problems found with final customized heuristics was higher as compared to the percentage of usability problems found with Nielsen's original set of heuristics. Also, from the Table V it is obvious that final customized heuristics finds more severe usability problems than the Nielsen's original set of heuristics. The average severity of usability problems found with Nielsen's set of heuristics is 2.8 and average severity of usability problems found with final customized heuristics is 3.06. According to Mankoff [12], the heuristics that finds many severe problems is more useful than a heuristic that finds fewer problems with lower severity. So, final customized heuristics is more useful in finding usability problems in human-humanoid robot interaction (HHRI) domain.

Figure 3 shows the increase in percentage of known usability problems found as the number of evaluators increases with Nielsen's original set of heuristics and final customized heuristics.

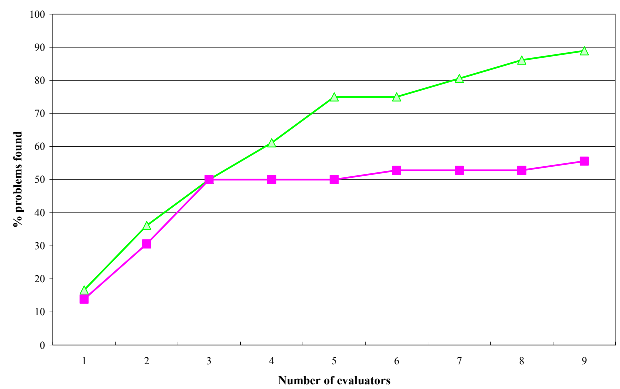

Fig. 3 Plot of Number of Evaluators Versus Percentage of Problem Found

But, it is clearly observable from the graph that the percentage of usability problems found with final customized heuristics is higher than the percentage of usability problems found with Nielsen's original set of heuristics for every group of evaluators experimented. The graph also shows that the final customized heuristics passed the canonical heuristic evaluation test as 3-5 evaluators has identified atleast 40-60% of the known usability problems as shown in Figure 3. A report detailing the identified usability problems was compiled and delivered to the development team as feedback for design improvement. The report clearly defined each of the usability problems with the severity rating. The human-humanoid robot interaction (HHRI) system of Robo-Erectus Junior was modified based on the inputs from the evaluators. From surveys and face to face interviews with humanoid robot handlers who control the robot during soccer matches, it was found that the modified interaction system exhibited better performance in terms of speed and flexibility in humanoid robot control during soccer matches as compared to the earlier design.

V. Conclusion

We have presented the development of customized heuristics for heuristic evaluation of human-humanoid robot interaction (HHRI) system in soccer domain. The customized heuristics were derived from the Nielsen's original set of heuristics. We have showed that the customized heuristics performs better than Nielsen's original set of heuristics for human-humanoid interaction (HHRI) domain. Our customized heuristics passed the canonical test with 3-5 evaluators finding 40-60% of known usability problems. We also found the estimated proportion of found problems compared to the theoretical total, and the estimated number of new problems expected to be found by including a specified number of new evaluators in the evaluation. Modified human-humanoid interaction (HHRI) system based on the feedbacks from the heuristic evaluation was found to perform better as compared to the earlier interaction system in terms of speed and flexibility in humanoid robot control during soccer matches. Potential future works in this direction includes, extension of the heuristics developed in this paper for multi-humanoid systems and humanoid-environment interactions.

References

[1] S. M. Metev and V. P. Veiko, "Task Structure and User Attributes as Elements of Human-Robot Interaction Design," in Proc. IEEE RO-MAN 06 Conference, Hatfield, United Kingdom, 2006.

[2] N. Weiss, and L. Hildebrand, "An exemplary robot soccer vision," CLAWAR/EURON Workshop on Robots in Entertainment, Leisure and Hobby, December 2004, Vienna, Austria.

[3] C.A. Calderon, C. Zhou, P.K. Yue, M. Wong, and M.R. Elara, "A distributed embedded control architecture for humanoid soccer robots," Proceeding of CLAWAR Conference 2007, Singapore.

[4] L. Hu, C. Zhou and Z. Sun, Estimating probability distribution with Q-learning for biped gait generation and optimization, in Proc. Of IEEE Int. Conf. on Intelligent Robots and Systems, 2006.

[5] J. Nielsen, "Finding usability problems through heuristic evaluation," in Proc. *ACM CHI'92 Conference,* Monterey, CA, 1992.

[6] J. Nielsen, *Heuristic evaluation*, J. Nielsen, and R. L. Mack, Eds.Wiley & Sons, New York, NY, 1994.

[7] J. Preece, *Interaction Design*, J. Preece, Y. Rogers, and H. Sharp, Eds. John Wiley & Sons, England, 2007.

[8] (2007) The Useit website. [Online]. Available: http://www.useit.com/papers/heuristic/heuristic_evaluation.html

[9] (2007) The Usabilitynet website. [Online]. Available: http://www.usabilitynet.org/tools/expertheuristic.htm

[10] C. Zhou and P.K. Yue, *Robo-Erectus: A low cost autonomous humanoid soccer robot*. Advanced Robotics, 18(7):717-720, 2004.

[11] (2007) Team Robo-Erectus website. [Online] Avaliable: http://www.robo-erectus.org.

[12] Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Ames, M., Lederer, S. "Heuristic evaluation of ambient displays," in Proc. CHI '03, pp. 169-176.