The USUS Evaluation Framework for Human-Robot Interaction

Astrid Weiss[1], Regina Bernhaupt[2], Michael Lankes[1] and Manfred Tscheligi[1]

Abstract. To improve the way humans are interacting with robots various factors have to be taken into account. An evaluation framework for Human-Robot Collaboration with humanoid robots addressing usability, social acceptance, user experience, and societal impact (abb. USUS) as evaluation factors is proposed (see figure 1). The theoretical framework USUS is based on a multi-level indicator model to operationalize the evaluation factors. Evaluation factors are described and split up into several indicators, which are extracted and justified by literature review. The theoretical factor indicator framework is then combined with a methodological framework consisting of a mix of methods derived and borrowed from various disciplines (HRI, HCI, psychology, and sociology). The proposed method mix allows addressing all factors within the USUS framework and lays a basis for understanding the interrelationship of the USUS Factors.

1 INTRODUCTION

Integrating humanoid robots into human working environments is a challenging endeavour. It is important to consider that users in human-robot interactions can face severe problems and difficulties. Studies showed that users perceive autonomous robots differently than other computer technologies [32]: Autonomous robots often lead to a far more anthropomorphic mental model than other interface technologies; moreover, as mobile robots always have to adapt to their environment they also have to conform to the humans they are working with, thus the interaction of robots and humans has to be negotiated. Furthermore, robots "learn" about themselves and their world, which heavily distinguishes them from traditional computing technology [32]. All these issues have a very strong influence on the users' work environment, on the way people collaborate, and on the way they experience robotic co-workers.

Such new technologies have a considerable impact on various factors of the interaction between humans and robots: usability, user experience, social acceptance, and societal impact have to be carefully investigated with appropriate measurements and approaches, to lay the basis for future ways of working, including robots that increase productivity and maintain safety. The theoretical and methodological evaluation framework USUS, which was developed from a human-centered HRI perspective [9] for evaluating usability, social acceptance, user experience, and societal impact for working scenarios with humanoid robots, can help us understand how to improve the construction of robots. The framework enables a positive user experience of all users – either individual or in groups, to enhance social acceptance and to support in general positive attitude towards (humanoid) robots in society.

[1] HCI&Usability Unit, ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. Email: {**firstname.lastname**}**@sbg.ac.at**.
[2] IHCS-IRIT, University Paul Sabtier, 118 Route de Narbonne, 31062 Toulouse Cedex 9. Email: regina.bernhaupt@irit.fr

2 STATE OF THE ART

As the research field of Human-Robot Interaction (HRI) is young, but evolving, the need for theoretical and methodological frameworks increases. As Bartneck et al. [4] claim: "If we are to make progress in this field then we must be able to compare the results from different studies". The framework proposed in this position paper should contribute to this aim and therefore take into account efforts already being made in this direction.

Thrun [43] provides one of the first theoretical frameworks for HRI based on the distinction of robots into three different kinds: industrial robots, professional service robots, and personal service robots. He describes in detail the different human-robot interface capabilities, different potential user groups, and the different contexts of use, which lays the first basis for future evaluation approaches in HRI. Thrun himself states in the abstract of this article: "The goal of this article is to introduce the reader to the rich and vibrant field of robotics, in hope of laying out an agenda for future research on human robot interaction" [43].

A similar intention was subject to Yanco et al. [48], who updated their taxonomy of human-robot interaction from 2002, to provide a basis for research in this area. They already address multiple research areas like HCI (Human-Computer Interaction), CSCW (Computer Supported Cooperative Work), and social sciences to offer a holistic picture of research aspects, proposing 11 categories, which need to be considered when investigating the interaction between a human and a robot. The taxonomy of Yanco et al. allows the comparison of different HRI research approaches and therefore is a first step in the direction of making HRI research more generalizable.

However, besides the theoretical frameworks, the need in HRI grows to define metrics to measure the success of robotic systems in a comparable way. This need became obvious with the inaugural workshop on "Metrics for Human-Robot Interaction" held in conjunction with the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008). The goal of this workshop was "to propose guidelines for the analysis of human-robot experiments and forward a handbook of metrics that would be acceptable to the HRI community and allow researchers both to evaluate their own work and to better assess the progress of others."

A first attempt in this direction was already made by Steinfeld et al. [42]. In their framework they proposed five metrics for task-oriented human-robot interaction with mobile robots: (1) Navigation, (2) Perception, (3) Management, (4) Manipulation, and (5) Social. Furthermore, they mention relevant “biasing effects”, which also have to be taken into consideration when evaluating the proposed metrics: communication factors (like delay, jitter, and bandwidth), robot response (like system lag and update rate), and the user (like training, motivation, and stress). Steinfeld et al. [42] provided above all metrics for usability factors (from an HCI perspective); however they stressed in their conclusion that their proposed “evaluation plan is to provide a living, comprehensive document that future research and development efforts can utilize as a HRI metric toolkit and reference source”.

Bartneck et al. [4] on the other hand tried to provide a standardized toolkit to measure user experience factors in HRI: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Based on an extensive literature review they developed five validated questionnaires using semantic differential scales to evaluate human-robot interaction in terms of these factors. “It is our hope that these questionnaires can be used by robot developers to monitor their progress”.

The Dutch research group around Marcel Heerink is focusing on the further development of the UTAUT model (Unified Theory of Acceptance and Use of Technology, see [46] for human-robot interaction in elder care institutions. They investigate which factors have an influence on the intention to use robotic agents. They consider factors like enjoyment [24] social presence [23] and social abilities [22]. Based on that, they want to develop a framework targeting studies on the acceptance of robotic agents by the elderly.

The theoretical and methodological framework proposed in this position paper addresses usability, social acceptance, user experience, and societal impact of humanoid robots used in collaborative tasks and intends to be applied to answer the general question ***if people experience robots as a support for cooperative work and accept them as part of society*** and thus give an holistic view on evaluating humanoid robots. Therefore, the proposed evaluation framework consists of two parts: (1) a theoretical framework defining the relevant evaluation factors and indicators combined with (2) a methodological framework explaining the methodological mix to address these factors during the evaluation of human-robot interaction.

3 THE FACTOR MODEL

The proposed evaluation framework for Human-Robot-Collaboration with humanoid robots is based on a multi-level indicator model targeting the factors *usability, social acceptance, user experience, and societal impact* as evaluation goals. The factors are selected to identify socially acceptable collaborative work scenarios where humanoid robots can be deployed beneficially to convince society to positively support the integration of humanoid robots in a human's working environment. The driving motivation for choosing these factors is to support user-centred evaluation approaches in HRI, going beyond pure usability studies. Although the framework was developed on an intense literature review taking into account existing frameworks and evaluation approaches in HRI, it cannot be guaranteed that is an exhaustive framework.

All factors chosen for evaluation are based on several indicators, which can be addressed with a methodological mix to be assessed during an iterative design process of human-robot collaboration. To justify the selection of factors and indicators several case studies are currently conducted in the framework of the EU-funded FP6 project “Robot@CWE: Advanced robotic systems in future collaborative working environments”. Figure 1 visualizes the combination of the theoretical and methodological framework.

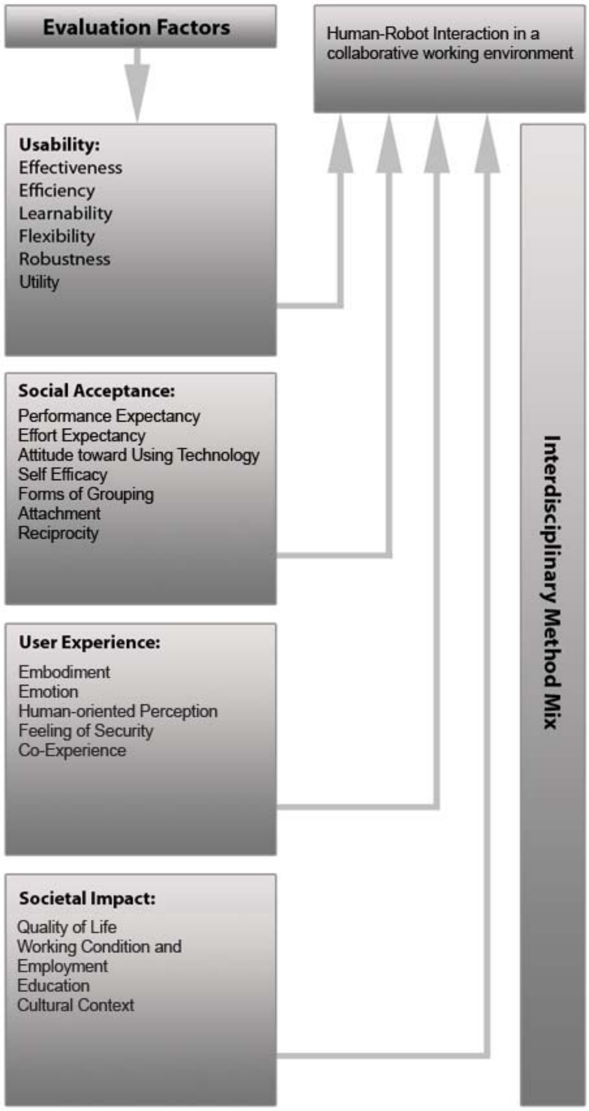

Figure 1: The Evaluation Framework

3.1 Usability as Evaluation Factor

The term usability refers to the ease of using an object. The iso924111:1998 [27] defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. This definition shows that usability is a concept of different indicators, than one single measurable term. Initial research on usability in human-robot interaction was mainly concentrating on indicators like, performance/ effectiveness and efficiency (e.g. [14], [40]). However, we propose that further aspects should be taken into account when assessing the usability of a humanoid robot.

3.1.1 Indicators for Usability

Effectiveness: The iso924111:1998 [27] defines effectiveness as "the accuracy and completeness with which users achieve specified tasks". Thus effectiveness describes how well a human-robot team accomplishes some task. This normally refers to the degree to which errors are avoided and tasks are carried out successfully solved, measured by e.g. "success rate" or "task completion rate".

Efficiency: In the iso924111:1998 [27] efficiency is defined as "the resources expended in relation to the accuracy and completeness with which users achieve goals". So efficiency is the rate or speed at which a robot can accurately and successfully assist humans.

Learnability: is one indicator of usability derived from software engineering. The concept of learnability is self-explanatory: how easy can a system be learned by novice users. This seems to be a key indicator for usability in human-robot interaction as robots are a technology people have almost no pre-experience with. Learnability incorporates several principles like familiarity, consistency, generalizability, predictability, and simplicity.

Flexibility: As humanoid robots are in general task independently designed (as they should be able to carry out a variety of tasks in unstructured environments and adapt to situations), flexibility seems to be another core-indicator for the usability evaluation of humanoid robots in collaborative working environments. Flexibility describes the number of possible ways how the user can communicate with the system.

Robustness: Novice users will produce errors when collaborating with humanoid robots, thus an efficient human-robot interaction has to allow the user to correct its faults on his/her own. Furthermore the robotic system itself should be error preventing, by means of being responsive and stable. Robustness is thus the level of support provided to the user to enable a successful achievement of tasks and goals.

Utility: As usability relates to the question of effectiveness and efficiency, like how well an interface supports the user to reach a certain goal or to perform a certain task, utility refers to how an interface can be used to reach a certain goal or to perform a certain task. The more tasks the interface is designed to perform, the more utility it has. Therefore utility and usability are related, but not interchangeable. Regarding humanoid robots utility is an essential factor, as a novice user has little knowledge about the utility of this type of robot as they are not designed for a specific task.

3.2 Social Acceptance as Evaluation Factor

Acceptance is an important issue to be evaluated in human-centred HRI. There is a need to find out the reasons why people accept robots in order to avoid rejection in a long term. Dillion [11] defines user acceptance as "the demonstrable willingness within a user group to employ technology for the tasks it is designed to support". However, the acceptance of autonomous acting robots cannot be defined that easily. In Western cultures a general retention of autonomous robots is present (e.g. [26], [30]) and furthermore novice users have difficulties in interpreting for which tasks a robot is designed. Thus, for socially situated robots [14], which can perceive and react on a social environment a different view on the term acceptance is necessary. Social acceptance within the USUS evaluation framework is defined as "an individual's willingness based on interaction experiences to integrate a robot into an everyday social environment". Several acceptance models exist which propose a theoretical framework for investigating technology acceptance, an excellent overview can be found in [45].

3.2.1 Indicators for Evaluating Social Acceptance

The indicators for the factor social acceptance in the USUS framework are derived from the UTAUT (Unified Theory of Acceptance and Use of Technology) model [46] (indicators 1 to 4) and from the theory of "object-centred sociality" [33] (indicators 5 to 8). The indicators are defined in accordance with the theory of object-centered sociality to understand the important aspect of how humans can be socially influenced in their working routines by a robot.

Performance Expectancy: In accordance to the UTAUT model Performance Expectancy is the strongest predictor of usage intention (it is significant at all points of measurement during the development of the model, both in voluntary and mandatory settings). "Performance expectancy is defined as the degree to which an individual believes that using the system will help him or her to attain gains in job performance." [46]..

Effort Expectancy: indicates to which extent the user perceives a system will be easy to use. Thus it includes believes of the degree of effort, difficulties, and understanding in usage, but also how complex users imagine the system to be. "Effort expectancy is defined as the degree of ease associated with the use of the system." [46].

Attitude toward Using Technology: In the UTAUT model the attitude toward using technology is defined as "an individual's overall affective reaction to using a system" [46]. In this evaluation framework attitude toward using technology is seen as the sum of all positive or negative feelings and attitudes about solving working tasks supported by a humanoid robot.

Self Efficacy: relates to a person's perception of their ability to reach a goal. This indicator is not included as a direct determinate in the UTAUT [46] but it is estimated to be a relevant factor for human-robot interaction. Perceived self efficacy can be defined as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Self Efficacy beliefs determine how people feel, think, motivate themselves and behave. Such beliefs produce these diverse effects through four major processes. They include cognitive, motivational, affective and selection processes" [3].

Forms of Grouping: Group practices are a core element of human behaviour. Grouping describes that humans who share certain characteristics, try to collect in a group. They interact preferably with other group members, share a common identity, and accept expectations and obligations of other group members. The question arising is whether humans can also share identity with robots.

Attachment: The term attachment was originally used to explain the bond that develops between a human infant and its caregiver [6]. In the last decades, the concept of emotional attachment has been used in a number of ways, also in relation to HRI (e.g. [29], [28]). Attachment may be defined as an affection-tie that one person forms between him/ herself and another person or object - a tie that binds them together in space and endures over time. According to Norman [38], emotional attachment can be seen as the sum of cumulated emotional episodes of users' experiences with a device in various context areas. These experience episodes can be categorized into three dimensions: a visceral level (first impression), a behavioural level (usage of the device), and a Reflective level (interpretation of the device).

Reciprocity: describes the principle of give-and-take in a relationship, but it can also mean the mutual exchange of performance and counter-performance. It is the positive or negative response of individuals towards the actions of others.

3.3 User Experience as Evaluation Factor

The term user experience is a very multifaceted concept and the research field of Human-Computer Interaction is still searching for a shared understanding of it [35]. A suitable definition for user experience of human-robot interaction can be adapted from Alben's general definition of user experience as “aspects of how people use an interactive product: the way it feels like in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it” [1].

Thus, users' experiences are related to a system and are embedded in a specific situation. Interaction goals, intra-psychological dispositions, the environment, involved people and the product itself have a significant impact on user experience [20].

3.3.1 Indicators for Evaluating User Experience

There is an increased interest in HRI establishing a positive experience for the interaction with a robot. In a working environment a positive user experience is desired since working tasks will be carried out more efficiently. In the following some factors of user experience in HRI are introduced, which are mainly derived from the framework of [15] to classify socially interactive robots.

Embodiment: describes the relationship between a system and its environment and can be measured by investigating the different perturbatory channels like morphology, which has impact on social expectations [15]. It is assumed that a humanoid form will ease the human-robot interaction because the rules for human social interaction will be invoked and thus a humanoid robot will provide a more intuitive interface, although this premise is largely untested. This is an assumption which is often found in literature (e.g. [7], [8]). Other researchers assume that robots are perceived as machines and that humanoid features therefore would generate unrealistic expectations or even fear [15].

Emotion: The indicator emotion implies that people tend to interact with computers (and robots) socially [39]. As emotion is an essential part in social interaction it has to be incorporated in the assessment and design of robots. Hassenzahl [19] structured aspects of user experience in product, goals related to the product, psychological status of a user, environment and people related to the experience. He stressed the importance of emotion in user experience by introducing relevant emotional episodes that are aroused during the interaction with a product. Users may experience satisfaction when a product fulfils the users' expectations. The emotion joy is felt by exceeding the user's expectations. Furthermore pride, surprise and attraction play a major role in experiencing a product.

Human-Oriented Perception: tries to simulate human perception. Social robots should be capable of tracking human features (e.g. face), interpreting human speech and recognize facial expressions etc. Robots should have different types of perception, meaning passive sensing and spoken language recognition. They have to be able to track people in different environment and lighting conditions. Additionally, the system should be able to recognize speech in two steps: speech processing and graph search. Furthermore vision based gesture recognition and facial perception (face detection and recognition) skills are necessary to guarantee Human-Oriented Perception. The system should not only be able to recognize facial display but also should have capabilities for communicating facial expression by means of image motion techniques, anatomical models, and principal component analysis.

Feeling of Security: As soon as humans should collaborate together with robots in the same environment, safety and security issues arise [10]. In addition to studies on how to eliminate the risk of hazards in human-robot collaboration [34], it is important to investigate how to design human-robot interaction in a way that humans experience them as safe. For example, [10] discovered that people prefer to be approached by a robot on the right hand side.

Co-Experience with Robots: Co-experience describes experiences with objects regarding how individuals develop their personal experience based on social interaction with others. People define “situations through an interpretive process in which they take into account the non-symbolic gestures and interpretations of others” [5]. Severinson-Eklund et al. e.g. [41] observed collaborative aspects of interaction with robots focusing on the personality of the robot, the communication paradigm between the user and the robot, and how a robot can mediate within a group of people.

3.4 Societal Impact as Evaluation Factor

The maturing of technology has been seen ever since as process that influences and changes society (e.g. consequences of the industrialization). Turn [44] conducted one of the first studies on the societal impact of computing technology. Turn pointed out that the purpose of societal impact studies should not only be to analyze the actual state of society, but to identify potential problems of future society, and to recommend corrective actions.

Societal impact can be defined as every effect of an activity on the social life of a community in general and more specific for the proposed framework: “Societal impact describes all effects the introduction of robotic agents consequences for the social life of a specific community (taking into account cultural differences) in terms of quality of life, working conditions and employment, and education.”

Theoretical assumptions on how the future society could look like and be influenced by robotic agents can above all be found in the cyborg and post-humanism literature (e.g. [2], [17]).

3.4.1 Indicators for Societal Impact

One of the main challenges when evaluating HRI is to predict the societal impact of robots. Already in 1981 the office of Technology Assessment conducted an exploratory workshop with the aim “to examine the state of robotics technology and possible public policy issues of interest”. Participants of this workshop identified four areas of “social issues” relevant for future societies in terms of the integration of robotic technology: (1) productivity and capital formation, (2) labor, (3) education and training, (4) international impact. Similar relevant impact factors could be derived from post-humanism and cyborg literature

Quality of Life, Health and Security: According to Gray [17] quality of human life is determined by several types of freedom, like free choice of gender orientations, or the freedom of travel. Furthermore, Gray argues that also stable human relationships, family constellations, and mutual reliance have important impact on human life quality and that the very nature of our relationships with each other will change through the integration of artificial intelligence into our environments.

Also the health system will be influenced by these developments, as high tech medicine will allow new therapy possibilities, which go hand in hand with the possibility of living longer. Security aspects could be concerned by the integration of intelligent robotic technology into everyday life, like electronic privacy, the freedom of consciousness, and the freedom of information. Already examples exist giving hints on these possible tendencies in future societies (e.g. [13], [16]). These researches efforts can on the on hand support the future health system and thus improve the quality of life, but on the other hand it could harm the natural relationships.

Working Conditions and Employment: Working conditions and employment includes all aspects affecting how people carry out their job and how employers take care of their employees, including things like working contracts, wages, working times, and work organization. Working conditions have always been affected by technology developments, as they can be used to increase the efficiency and productivity. This in turn may lead to an increasing degree of replacement of e.g. assembly-line workers by robots, as robots can complete some physical tasks much quicker and more precisely than humans, e.g. harvesting [18]. Forlizzi et al. e.g. think of closing the gap of lacking service personal in hospitals by the introduction of care robots, for situations when no physical presence of a doctor or a nurse is necessary [16].

Education: New software, new sciences and new disciplines require new types of education. Lifelong learning is necessary to manage duties and responsibilities. To avoid the fear of being displaced by a robot it might be necessary to give educational advertising. In times of increasing utilization of robots the aspect of education should not be disregarded. Considering the situation now, after the launch of computers, every child in western society is taught how to use computers and how to use certain programs. But without this education most people would not be prepared sufficiently for the labour market. So it might be necessary to be prepared for utilization of robots, in the physical manner, but potentially in a psychological manner too.

Cultural Context: Culture embraces the whole range of practices, customs and representations of a society. In their rituals, stories and images, societies identify what they perceive as good and evil, proper, and racially different [2]. However, culture does not exist in the abstract. On the contrary, it is in the broadest sense of the term textual, inscribed in the paintings, operas, sculptures, furnishings, fashions, bus tickets and shopping lists which are the currency of both aesthetic and everyday exchange [2]. Thus the socio-cultural environment plays a decisive role. Japanese or South Koreans interact with robots in a quite different and more enthusiastic manner than in Europe, where people are more sceptical. This is due to the fact that in Japan e.g. automatons have a long tradition in religious ceremonials. Furthermore, in special Japanese religions a soul to things and machines is granted. Last but not least the positive presentation of robots in Japanese literature leads to a high acceptance too. There are also great differences in Japan and Europe regarding robots in the working area. Japanese employees trust in the corporation's decision and establish a long and quasi familiar relationship to their corporation. In contrast, in Europe due to short employment contracts and numerous structural changes over the last years robots are often perceived as a rationalization instrument [26].

4 THE METHODOLOGICAL FRAMEWORK

Different methods are available to assess interactive systems in general and human-robot collaboration in specific. Depending on the data that is required, the resources that are available (i.e., time and money) and the design phase, a decision can be made which methods to choose [12]. One of the basic principles of a user-centred evaluation approach is that potential users evaluate whether the system is usable, acceptable etc. or not. Moreover, the method selected has to be adapted to the context, the tasks and the system which will be evaluated. Formative evaluation approaches are the main interest of this framework, as summative evaluations in real workplace environments with humanoids are hardly possible.

To investigate and evaluate the above defined indicators for usability, social acceptance, user experience and societal impact, a combination of different methods is needed. Qualitative research is combined with quantitative measures for the evaluation approach. The proposed methods, presented in a matrix-diagram visualizing which method is suitable for which indicator, are presented later on in more detail.

Table 1: The Methodological Mix

	Methods	Expert Eval	User Studies	Questionnaires	Physio. Measures	Focus Groups	Interviews
Research Objectives							
Usability							
	Effectiveness	X	X				
	Efficiency	X	X				
	Learnability	X	X				
	Flexibility	X	X				
	Robustness	X	X				
	Utility			X			X
Social Acceptance							
	Performance Expectancy			X		X	
	Effort Expectancy			X		X	
	Attitude toward Using Technology			X			
	Self Efficacy			X		X	
	Forms of Grouping			X		X	
	Attachment			X		X	
	Reciprocity			X			
User Experience							
	Embodiment			X		X	
	Emotion			X	X	X	
	Human-Oriented Perception			X			
	Feeling of Security			X	X	X	
	Co-Experience			X		X	
Societal Impact							
	Quality of Life			X		X	X
	Working Conditions			X		X	X
	Education			X		X	X
	Cultural Context			X		X	X

4.1 Expert Evaluation

In traditional HCI research expert evaluations are used to asses a system in terms of its usability and detect as many usability problems as possible in a way that is less cost- and effort-intensive than user testing.

Heuristic Evaluation: A heuristic evaluation is intended to find and describe usability problems of a system on the basis of fundamental principles, so called heuristics [37]. Heuristics are describing essential attributes that a system should feature to ensure that the user is able to perform a task within a specified context in an effective, efficient, and satisfying way. Such a heuristic evaluation is usually performed by a small team of interface experts inspecting the system and comparing to what extent the principles have been adopted. All experts then have to rank all problems according to their severity. Thus, the result of a heuristic evaluation is a complete list of all detected usability problems ranked according to their severity.

Cognitive Walkthrough: A cognitive walkthrough [47] is conducted by at least two usability experts assessing the usability of a system based on predefined task structures. The expert evaluators try to imagine how a typical (potential) user would solve a task with the assumption of minimizing the cognitive load. Thus, the cognitive walkthrough is above all used to evaluate the usability of a system in terms of its learnability and how intuitive it can be used. During each task analysis the experts ask themselves a set of questions for each subtask.

4.2 User Studies

Laboratory-based: User studies are used to provide empirical evidence to answer a concrete research question or hypothesis. Such user-involving evaluations are based on tasks, which subjects conduct, while their behaviour is observed and measured by a researcher. Classical metrics measured during user studies are task completion rate or error rate to measure the effectiveness of the tested system, and task duration to measure the efficiency. To better understand usability problems of the subjects occurring during the solving of task the "think aloud" method is often applied in user studies. "Think aloud" means that subjects are asked to say whatever they are looking at, thinking of, doing, and feeling, as they conduct a task. This enables observers to see first-hand protocol usability problems. User studies are normally audio and video taped so that researchers can go back and refer to what subjects did, and how they reacted. User studies provide a good opportunity to be combined with other methods like questionnaires or qualitative interviews.

Field-based: Field based user studies have the focus of interest in testing the usage of a system in a realistic usage context. This kind of user studies can, but does not necessarily have to be task-based. In general the procedure is similar to user studies conducted in the laboratory, whereas the observation is mostly passive and unstructured. However, as field trials have to take into account more disturbing factors (e.g background noise or light conditions), the interpretation and analysis of the data, is more difficult.

Wizard of OZ Technique: User studies can be conducted with fully autonomous systems, or they can be based on the so-called "Wizard of OZ Technique", short "WOZ". [31]. To allow user testing in very early stages of the prototype development, which cannot be fully implemented at that stage, a human "Wizard" enacts (or simulates) the system features in interaction. This approach integrates several advantages for user studies in HRI, as e.g. safety and security issues can be controlled during the testing, and relevant social cues can be simulated. However, also disadvantages are incorporated when evaluating user experience and social acceptance: Is the perception of the robot measured, or the perception of a human wizard...?

4.3 Standardized Questionnaires

A questionnaire is a research instrument that consists of a series of questions with the purpose of gathering statistically analyzable data from the participants. Standardized questionnaires are based on closed answers, where participants have only to choose one of the pre-defined answers making it easier for them to complete the questionnaire. A good overview on existing questionnaires for HRI research can be found in [2].

4.3 Physiological Measurements

Physiological measurements can give valuable additional input to e.g. questionnaires and focus groups and can detect information participants for instance do not want to state. Several methods can be used to measure the emotional state of a subject [36]: Methods like this can identify "emotions" at exactly the time they happen during a user study and can be combined with e.g. reflective questionnaire data on the emotional state. This can support investigation of user experience factors in terms of "heteronom and autonom identification".

4.4 Focus Groups

Focus groups allow the researcher to explore participants' attitudes, beliefs, and desires in great depth and give insights how they experience a system. Focus groups are structured discussions about specific topics and moderated by a trained leader. The focus of the discussion is also triggered by the participants' selection which is based on common characteristics, as opposed to differences among them. It is important to note that a focus group only gathers qualitative data, which can be used as input to further develop other research instruments which gather quantitative and generalizable results.

4.5 In-depth Interviews

In-depth interviews are a qualitative research technique, which allows "person-to-person" discussion on a specific topic. It aims to get increased insight participant's ideas, attitudes, and feelings on the discussed issues. In-depth interviews can be combined with user studies to discuss with the participants how they experienced the interaction with the robotic system.

Expert Interviews: This specific type of in depth interview is conducted in the same way as qualitative interviews in general. However, they are conducted with people who are considered experts in a particular subject, e.g. humanoid robot development, robots in the workplace, robots and ethics etc. The aim is not to get to know their attitude or feelings on a topic, but that they share their knowledge.

Delphi study: is a special form of how to conduct expert interviews. The goal is to find a solution for a complex problem statement. Delphi studies consist of several rounds of interviews and discussions with several experts. After each round researchers report to experts the results of the previous discussion round as group opinion about the problem statement. This process happens as long as a common solution for the problem statement is found.

5 SUMMARY AND OUTLOOK

The goal of the framework is a multi-level evaluation model covering a multitude of factors: Usability, social acceptance, user experience, and societal impact. These four factors are called the USUS factors. Main goal of the evaluation framework is to guide research in answering the question on how people experience robots as a support for collaborative work and accept them as part of society. The framework operationalize the USUS factors with indicators and describes methodological possibilities to investigate them - and beyond that, helps to The evaluation framework is intended to guide current activities within the Robot@CWE research project, but it can help other researchers on a more general level to understand what kind of methods can be helpful to investigate the USUS factors in human-robot interaction. To assess the validity of the proposed framework currently more than twenty evaluations, at seven different sites involving various types of humanoid robots, are conducted to investigate the relationship between the various factors (first results can be found in ([49][50][51]). Further results will be available in 2009, showing advantages and limitations of the proposed framework to evaluate human-robot collaboration with humanoid robots, but we expect this framework to be a valuable help for researchers to investigate the USUS factors in HRI.

ACKKNOWLEDGEMENTS

This framework was developed in the framework of the ROBOT@CWE project (funded by FP6-2005- IST-5). The authors like to thank all partners from the project consortium.

REFERENCES

[1] L. Alben, 'Quality of experience: defining the criteria for effective interaction design', interactions, 3(3), 11–15, (1996).

[2] N. Badmington, Posthumanism, Readers in cultural criticism, Palgrave, Houndmills, Basingstoke, Hampshire , New York, 2000.

[3] A. Bandura, Self-efficacy: The exercise of control, New York, NY, Freeman, 1997.

[4] C. Bartneck, E. Croft, and D. Kulic, 'Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots', International Journal of Social Robotics, (2009).

[5] K. Battarbee, 'Defining co-experience', in DPPI '03: Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp. 109–113, New York, NY, USA, (2003). ACM.

[6] J. Bowlby, 'The nature of the child's tie to his mother', *International Journal of Psychoanalysis*, **39**, 350–373, (1958).

[7] C. Breazeal and B. Scassellati, 'How to build robots that make friends and influence people', in *IROS '99: IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pp. 858–863, (1999).

[8] R. Brooks, 'Humanoid robots', *Commun. ACM*, **45**(3), 33–38, (2002).

[9] K. Dautenhahn, 'The art of designing socially intelligent agents : Science, fiction, and the human in the loop', *Applied Artificial Intelligence*, **12**(7), 573 – 617, (1998).

[10] K. Dautenhahn, M. L. Walters, S. Woods, K. L. Koay, E.A. Nehaniv, C. L. Sisbot, R. Alami, and T. Simeon, 'How may I serve you?: a robot companion approaching a seated person in a helping context', in *HRI '06: Conference on Human-Robot Interaction*, pp. 172–179, Salt Lake City, Utah, USA, (3 2006). ACM.

[11] A. Dillon, 'User acceptance of information technology', in *Encyclopedia of Human Factors and Ergonomics*, ed., W. Karwowski, Taylor and Francis, London, (1 2001).

[12] A. Dix, G. D. Abowd, and J. E. Finlay, *Human-Computer Interaction (3rd Edition)*, Prentice Hall, 2004.
[13] R. Dobson, 'Meet Rudy, the world's first "robodoc"', *BMJ*, **329**(7464), 474, (2004). 1 University of Leipzig, Germany, email: somename@informatik.unileipzig. de
[14] J. L. Drury, J. Scholtz, and H. A. Yanco, 'Awareness in human-robot interactions', in *IEEE Conference on Systems, Man and Cybernetics*, pp. 912– 918, Washington, (10 2003).
[15] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143–166, (3 2003).
[16] J. Forlizzi, 'Robotic products to assist the aging population', *interactions*, **12**(2), 16–18, (2005).
[17] C. H. Gray, Cyborg citizen: Politics in the posthuman age, Routledge, New York, 2001.
[18] S. Green, X. Q. Billinghurst, M.and Chen, and G. Chase, 'Human-robot collaboration: A literature review and augmented reality approach in design', *International Journal of Advanced Robotic Systems*, **5**(1), 1– 18, (2008).
[19] M. Hassenzahl, 'The effect of perceived hedonic quality on product appealingness', *Int. J. Hum. Comput. Interaction*, **13**(4), 481–499, (2001).
[20] M. Hassenzahl, 'The thing and i: understanding the relationship between user and product', in *Funology. From Usability to Enjoyment*, eds., M. Blythe, C. Overbeeke, A. F. Monk, and P. C. Wright, 31–42, Kluwer, Dordrecht, (2003).
[21] M. Heerink, and B. Kröse, B. Wielinga, and V. Evers, 'Studying the acceptance of a robotic agent by elderly users', *International Journal of Assistive Robotics and Mechatronics*, **7**(3), 25–35, (9 2006).
[22] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of a robot's social abilities on acceptance by elderly users', in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 521–526, (2006).
[23] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users', in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pp. 695–700, (2008).
[24] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, 'Enjoyment intention to use and actual use of a conversational robot by elderly people', in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 113–120, New York, NY, USA, (2008). ACM.
[25] P. J. Hinds, T. L. Roberts, and H. Jones, 'Whose job is it anyway? a study of human-robot interaction in a collaborative task', *Human-Computer Interaction*, **19**(1-2), 151–181, (2004).
[26] T. N. Hornyak, *Loving the machine: the art and science of Japanese robots*, Kodansha international, Tokyo, New York, London, 2006.
[27] ISO 9241-11, Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability, International Organization for Standardization, 1998.
[28] F. Kaplan, 'Free creatures: The role of uselessness in the design of artificial pets', in *Proceedings of the 1st Edutainment Robotics Workshop*, (September 2000).
[29] F. Kaplan, 'Artificial attachment: Will a robot ever pass the ainsworth's strange situation test?', in *Proceedings of Humanoids 2001: IEEE-RAS International Conference on Humanoid Robots*, pp. 125 – 132, (2001).
[30] F. Kaplan, 'Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots', *International journal of humanoid robotics*, **1**(3), 465–480, (2004).
[31] J. F. Kelley, 'An iterative design methodology for user-friendly natural language office information applications', *ACM Trans. Inf. Syst.*, **2**(1), 26–41, (1984).
[32] S. Kiesler and P.J. Hinds, 'Introduction to this special section on human-robot interaction', *Human-Computer Interaction, Spec. Issue on Human-Robot Interaction*, (1-2), (2005).
[33] K. Knorr-Cetina, 'Sociality with objects: Social relations in postsocial knowledge societies', *Theory Culture Society*, **14**(4), 1–30, (November 1997).
[34] D. Kulic and E. Croft, 'Strategies for safety in human robot interaction', in *ICAR'03: IEEE International Conference on Advanced Robotics*, pp. 644–649, Coimbra, Portugal, (2003).
[35] E. Law, V. Roto, A. Vermeeren, J. Kort, and M. Hassenzahl, 'Towards a shared definition of user experience', in *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pp. 2395–2398, New York, NY, USA, (2008). ACM.
[36] M. Minge, Methoden zur Erhebung emotionaler Aspekte bei der Interaktion mit technischen Systemen, Ph.D. dissertation, Freie Universitaet Berlin, Berlin, 2005.
[37] J. Nielsen, 'Finding usability problems through heuristic evaluation', in *Proceedings of the ACM CHI 92 Human Factors in Computing Systems Conference*, pp. 373–380. ACM Press, (1992).
[38] D. A. Norman, Emotional Design: Why We Love (Or Hate) Everyday Things, Basic Books, New York, USA, 2004
[39] B. Reeves and C. Nass, The Media Equation: How People Treat Computers, Televisions, and New Media Like Real People and Places, Cambridge University Press, New York, 1996.
[40] J. Scholtz, 'Evaluation methods for human-system performance of intelligent systems', in *PerMIS'02: Performance Metrics for Intelligent Systems Workshop*, Gaithersburg, MD, USA, (8 2002).
[41] K. Severinson-Eklundh, Anders Green, and Helge Hüttenrauch, 'Social and collaborative aspects of interaction with a service robot', *Robotics and Autonomous Systems*, **42**(3-4), 223–234, (2003).
[42] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, 'Common metrics for human-robot interaction', in *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 33–40, New York, NY, USA, (2006). ACM.
[43] S. Thrun, 'Toward a framework for human-robot interaction', *Human-Computer Interaction*, **19**(1), 9–24, (2004).
[44] R. Turn, 'Courses on societal impacts of computers', *SIGCAS Comput. Soc.*, **13, 14**(4, 1-3), 14–16, (1984).
[45] V. Venkatesh and F. D. Davis, 'A theoretical extension of the technology acceptance model: Four longitudinal field studies', *Manage. Sci.*, **46**(2), 186–204, (2000).
[46] V. Venkatesh, M. G. Morris, G. B. Davi, and F. D. Davis, 'User acceptance of information technology: Toward a unified view', *MIS Quarterly*, **27**(3), (2003).
[47] C. Wharton, J. Bradford, R.Jeffries, and M. Franzke, 'Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations', in *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 381–388, New York, NY, USA, (1992). ACM.
[48] H.A. Yanco and J. Drury, 'Classifying human-robot interaction: an updated taxonomy', *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, **3**, 2841–2846, (Oct. 2004).
[49] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kühnlenz, and M. Buss. Methodological variation for acceptance evaluation of human-robot interaction in public places, in *IEEE RO-MAN 2008: Proceedings of the International Symposium on Robot and Human Interactive Communication,* Munich, Germany, (2008).
[50] A. Weiss, R. Bernhaupt, M. Tscheligi, and E. Yoshida. 'Addressing user experience and societal impact in a user study with a humanoid robot', in *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction (accepted for publication)*, (2009).
[51] A. Weiss, D. Wurhofer, M. Lankes, and M. Tscheligi. 'Autonomous vs. tele-operated: How people perceive human-robot collaboration with HRP-2', *poster presentation at HRI 2009:ACM/IEEE International Conference on Human-Robot Interaction*, (2009).