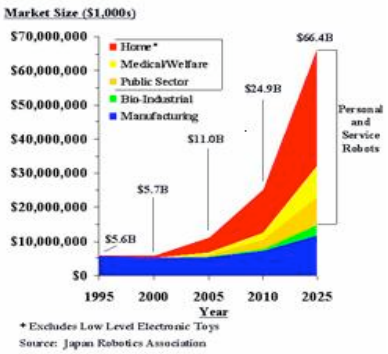
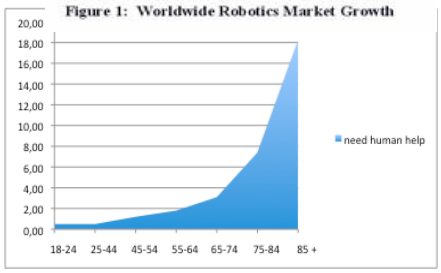
Common metrics for HRI

Mohamed CHETOUANI
mohamed.chetouani@upmc.fr
Institut des Systèmes Intelligents et de Robotique (ISIR)
UMR 7222

Introduction

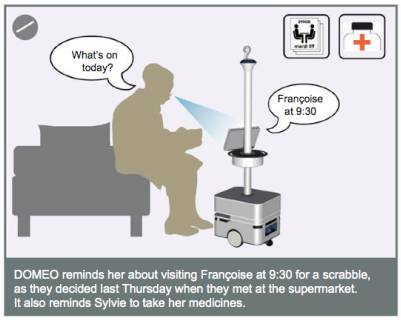
- The number of personal and service robots sold is expected to increase
- Domestic robots user presents a unique challenge





Introduction

- Elderly users are likely to be uncomfortable with domestic robots due to a lack of exposure to technology, disabled users might have difficulties using robots that do not provide **interaction modalities that accommodate their needs**



DOMEO reminds her about visiting Françoise at 9:30 for a scrabble, as they decided last Thursday when they met at the supermarket. It also reminds Sylvie to take her medicines.

robosoft Advanced Robotics Solutions

3

Introduction

- The evaluation of interactions between robots and humans, different types of interactions, and individual robot and human behaviors requires adequate metrics and guidelines
- Metrics are not comparable due to bias towards application specific measures
- What to evaluate?
  - **Behaviors**
  - **Ways of interactions**
  - **Social and psychological aspects**
  - **Technical characteristics and objective measures: success rates, interaction time, error rates…**

4

Introduction

- Why evaluation is important?
  - **Comparison of HRI systems, protocols, robots (embodiment), social acceptance**
  - **Feasibility, Efficiency, Safety**
  - **Impact on the user: rehabilitation, usability, societal impact**
  - **User experience**
  - **To improve the way Humans are interacting with robots**
- What to evaluate?
  - **Behaviors**
  - **Ways of interactions**
  - **Social and psychological aspects**
  - **Technical characteristics and objective measures: success rates, interaction time, error rates…**

5

Introduction

- Methodology:
  - Large sample sizes of participants: statistical analyses, quantitative methods, large-scale experiments, population/robot modes comparisons
  - Case studies: single-case analysis, qualitative approaches
  - Longitudinal analyses: repeated measurement of subjects over time
- Most studies cannot be reproduced…

6

Metrics

- Task effectiveness: How well a Human-Robot team accomplishes some task?
- Neglect tolerance: Autonomy of a robot with respect to some task
- Robot attention demand: How much a robot is demanding?
- Fan-out: Number of robots simultaneously operating

Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

7

Metrics

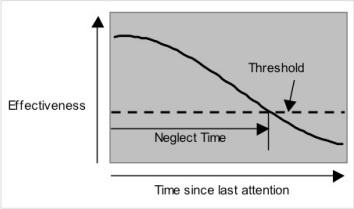
- Task effectiveness: How well a Human-Robot team accomplishes some task?
  - Time-based metrics: speed of performance
  - Error metrics: mistakes, damage
  - …
- Metrics are task specific
- Overall task effectiveness is best measured after the task is complete

Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

8

Metrics

- Neglect tolerance (NT):
  - Measures the autonomy of a robot with respect to the task
  - Measures of how robot's effectiveness declines over time when the robot is neglected by the user

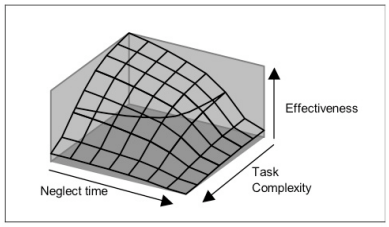


Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

9

Metrics

- Neglect tolerance (NT):
  - Function of task complexity:



  - Measuring neglect tolerance: Amount of time that a Human can ignore a robot
    - *"Place the robot at a some random location in a problem world and giving it a random goal to achieve and measure the time that the robot is effective: the elapsed time during which the robot makes progress towards that goal before dropping below the effectiveness threshold"*
    - *"Time between some user instruction and either dropping below the effectiveness threshold or a new user instruction"*

Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

10

Metrics

- Robot Attention Demand:
  - How much a robot is demanding (RAD)…
  - Measure of the fraction of total time that a user must attend to a given robot
    - Interaction effort (IE): can be the amount of time required to interact with the robot

Amount effort that the user must expend interacting

$$RAD = \frac{IE}{IE + NT}$$

  - RAD is unitless quantity that represents the fraction of a human's time that is consumed by interacting with a robot

- Examples:
  - The teleoperated robots have a small NT and RAD approaches to 1: the user can focus on other things besides interacting with the robot
  - Reducing RAD can be done by increasing NT or decreasing IE but not always (NT ang IE are NOT independent)

Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

11

Metrics

- Robot Attention Demand:
  - From RAD, you can also define the user's free time:

$$FT = 1 - RAD$$

  - Fraction of the task tome that the user does not need to pay attention to the robot
  - It can also be used to measure RAD
    - One way it to give to the user a robotic task and a secondary task

Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

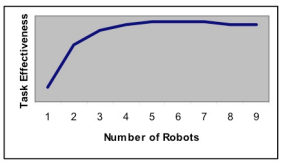
12

Metrics

- Fan-Out :
  - One way to leverage Human attention is to allow a user to operate multiple robots simultaneously:
    - To accomplish some tasks more quickly and effectively: surveillance, exploration…
  - Measure of the effectiveness of a Human-Robots team using FAN-OUT
  - Fan-Out is an estimate of the number of robots that a user effectively operates at once

$$FO = \frac{1}{RAD} = \frac{IE + NT}{IE}$$

  - FO increases as NT becomes large relative to interaction effort
  - Task effectiveness increases as more robots are added to the task



Olsen and Goodrich, Metrics for evaluating Human-Robot Interaction, Proc. NIST Performance Metrics for Intelligent Systems Workshop, 2003

Common metrics

- Steinfeld et al proposed 5 metrics for task-oriented Human-Robot Interaction with mobile robots:
  - Navigation
  - Perception
  - Management
  - Manipulation
  - Social

Steinfeld et al., Common Metrics fo Human-Robot Interaction, HRI 2006

USUS evaluation framework

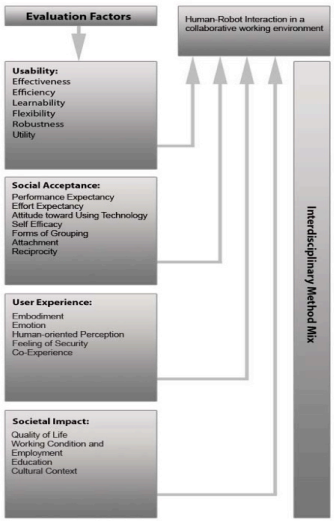
- New technologies have considerable impact on various factors of the interaction between Humans and robots:
  - usability, user experience, social acceptance and social impact
- These factors have to be investigated with appropriate measurements and approches
- This is the purpose of the theoretical and methodological evaluation framework USUS:
  - Usability, Social acceptance, User experience and Social impact

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

15

USUS evaluation framework

- Human-Robot collaboration framework
- Multi-level indicator model
- Factors are selected to identify socially acceptable collaborative work scenarios where humanoids can be deployed beneficially
- User-centred evaluation approach



Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

USUS evaluation framework

- Usability
- The term usability refers to the ease of using object
  - ISO924111:1998 defines usability as:
    - "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction"
- Usability is a concept of different indicators han one singel measurable term
- In HRI, it is usually measured as performance/effectiveness and efficiency

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

17

USUS evaluation framework

- Indicators for usability
  - Effectiveness: "the accuracy and competences with which users achieve specified tasks" => *success rate or task completion rate*
  - Efficiency: "the ressources expended in relation to the accuracy and completeness with which users achieve goals" => *rate or speed at which a robot can accurately and successfully assist Humans*
  - Learnability: "How easy can a system be learned by novice users?" => *familiarity, consistency, predictability, simplicity*
  - Flexibility: "describes the number of possible ways how the user can communicate with the system"
  - Robustness: Novice users will produce errors when collaborating with robots, thus an efficient HRI system has to allow the usr to correct its faults on his/her own. => *error preventing, responsive and stable*
  - Utility: "How an interface can be used to reach a certain goal or to perform a certain task?"

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

18

USUS evaluation framework

- Social acceptance
- There is a need to find out the reasons why people accept robots in order to avoid rejection in a long term
- "The demonstrable willingness within a user group to employ technology for the tasks it is designed to support"
- Acceptance is culture dependent
- It depends on the user experience
- USUS framework defines social acceptances as "an individual's willingness based on interaction experiences to integrate a robot into an everyday social environment"

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

19

USUS evaluation framework

- Indicators for evaluating Social Acceptance
  - Performance expectancy: defined as the degree to which an individual believes that using the system will help him or her to attain gains in job performance.
  - Effort expectancy: defined as the degree of ease associated with the use of the system. Extent the user perceives a system will be easy to use.
  - Attitude toward using technology: defined as an individual's overall affective reaction to using a system. Sum of all positive or negative feelings and attitudes about solving working tasks supported by a robot.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

20

USUS evaluation framework

- Indicators for evaluating Social Acceptance
  - Self efficacy: relates to a person's perception of their ability to reach a goal.
  - Forms of grouping: The question arising is whethever Humans can also share identity with robots.
  - Attachment: The term was originally used yo explain the bond develops between a Human infant and its caregiver. Emotional attachment can be seen as the sum of cumulated emotional episoded of users' experiences with a device un various context areas.
  - Reciprocity: describes the principle of give-and-take in relationship, but in can also mean the mutual exhange of performance and counter-performance.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

21

USUS evaluation framework

- User experience
  - *"Aspects of how people use an interactive product:*
    - *The way it feels like in their hands,*
    - *How they understand how it works,*
    - *How they feel about it while they're using it,*
    - *How well it serves their purposes*
    - *How well it fits into the entire context in which they are using it."*
- Users' experience are related to a system and are embessed in a specific situation.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

22

USUS evaluation framework

- Indicator for evaluating User experience
  - Embodiment: describes the relationship between a system and its environment and can be measured by investigating the different pertubatory channels like morphology, which has impact on social expectations.
  - Emotion: The indicator emotion implies that people tend to interact with robots socially. Users may experience satisfaction when a product fulfils the users' expectations.
  - Human-oriented perception: tries to simulate Human perception by tracking human features, interpreting speech… but also communicating facial expresions

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

23

USUS evaluation framework

- Indicator for evaluating User experience
  - Feeling of security: As soon as Humans collaborate together with robots in the same environment, safety and security issues arise. => Dautenhahn et al discovered that people prefer to be approached by a robot on the right hand side.
  - Co-experience with robots: How individuals develop their personal experience based on social interaction with others.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

24

USUS evaluation framework

- Social impact
  - *"describes all effects the introduction of robotic agent consequences for the social life of a specific community (raking into account cultural differences) in terms of quality of life, working conditions and employment, and education"*

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

25

USUS evaluation framework

- Indicators for Social impact
  - Quality of life, health and security: new therapy possibilities, security aspects…
  - Working conditions and employment: includes all aspects affecting how people carry out their job and how employers tak care of their employees (working times, organization…).
  - Education: New software, new sciences and new disciplines require new types of education. Preparation for utilization of robots, in the physical manner and in psychological manner too.
  - Cultural context: whole range of practices, customs and representations of a society. Example: Japanese or South Koreans interact with robots in quite different manner tha in Europe.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

26

Methodological framework

- User-centred evaluation approach
- Qualitative research is combined with quantitative measures for the evaluation approach.

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

Methodological framework

- Expert evaluation:
  - Heuristic evaluation: Find and describe usability problems on the basis of fundamental principles. The result is a list of all detected usability problems ranked according to their severity.
  - Cognitive walktrough: is conducted by at least 2 usability experts assessing the usability of a system based on predifined task structures. The expert evaluators try to imagine how a typical (potential) user would solve a task with the asssumption of minimizing the cognitive load.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

Methodological framework

- User studies:
  - Laboratory-based: User studies are used to provide empirical evidence to answer a conrete research question or hypothesis. This enables observers to see first-hand protocol usability problems. User studies are normally audio and video tapes so that researchers can go back and refer to what subjects did, and how they reacted.
  - Field-based user studies: have the focus of interest in testing the usage of a system in a realistic usage context. Take into account more disturbbing factors (e.g. background noise, light conditions…), the interpretation and analysis of the data is more difficult.
  - Wizard of OZ: To allow user testing in very early stages of the prototype. A Human "wizard" simulates the system features in interaction:
    - Advantages: safety, security, relevant social cues
    - Drawbacks: Is the perception of the robot measured or the perception of a Human "wizard"?

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

29

Methodological framework

- Standardized questionnaires:
  - A questionnaire is research instrument that consists of a series of questions with the purpose of gathering statistically analyzable data from the participants
- Physiological measurements:
  - They can give valuable additional input to other evaluation methods (questionnaires…).
  - Detection of "intrinsic" users' states.
- Focus groups:
  - They allow the researcher to explore participants" attitudes, beliefs, and desires in great depth.
  - Focus groups are structured discussion about specific topics and moderated by a trained leader.
- In-depth interviews:
  - Expert interviews: "person-to-person" discussion, share knowledge
  - Delphi studies: Find a solution for a complex problem statement. Discussions with several experts.

Weiss et al., The USUS Evaluation Framework for Human-Robot Interaction, New Frontiers in Human-Robot Interaction, 2009

30

Debugging HRI systems

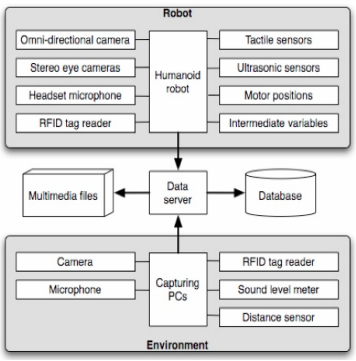
- Controlled experiments and field trials are regularly conducted to stage HRI.
- Using User-centred metrics
- But how to debug HRI systems?

Kooijmans et al., Interaction debuggin: an Integral Approach to Analyze Human-Robot Interaction , HRI 2006

31

Debugging HRI systems

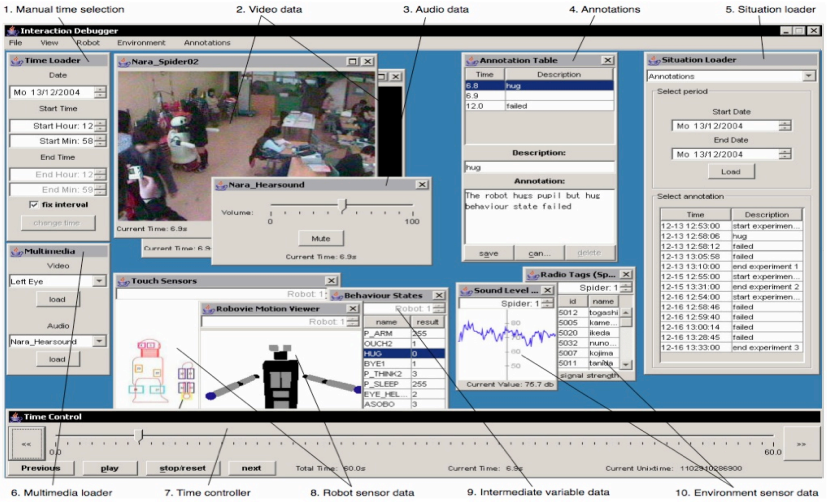
- Interaction debugging approach:
  - Collection of multi-modal data
  - Data analysis



Kooijmans et al., Interaction debuggin: an Integral Approach to Analyze Human-Robot Interaction , HRI 2006

32

Debugging HRI systems



Kooijmans et al., Interaction debuggin: an Integral Approach to Analyze Human-Robot Interaction , HRI 2006

33

Example of RobotCup@Home

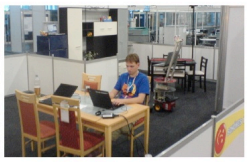
This is a list of the current desired technical abilities which the tests in RoboCup@Home will focus on.

- Navigation in dynamic environments
- Fast and easy calibration and setup
  The ultimate goal is to have a robot up and running out of the box.
- Object Recognition
- Object Manipulation
  Manipulation is essential for almost any future home applications.
- Recognition of Humans
- Human Robot Interaction
  An aim of the competition is to foster natural interaction with the robot using speech and gesture commands.
- Speech recognition
  For intuitive interaction it is essential to come up with solutions that do not require headsets in the future.
- Gesture recognition
- Robot applications
  RoboCup@Home is aiming for applications of robots in daily life.
- Ambient intelligence
  Communicate with surrounding devices, getting information from the the Internet, e.g. Asking the robot about the weather, reading/writing emails.

34

Example of RobotCup@Home

(a) RoboCup 2006 Bremen - living room



(b) RoboCup 2006 Bremen - kitchen area

(c) RoboCup German Open 2007 Hannover

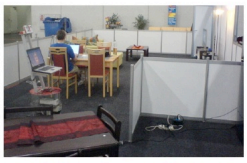
(d) RoboCup 2007 Atlanta

(e) RoboCup German Open 2008 Hannover

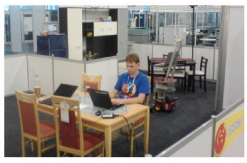
(f) RoboCup 2008 Suzhou

35

Example of RobotCup@Home



(a) RoboCup 2006 Bremen - living room



(b) RoboCup 2006 Bremen - kitchen area

(c) RoboCup German Open 2007 Hannover

(d) RoboCup 2007 Atlanta

(e) RoboCup German Open 2008 Hannover

(f) RoboCup 2008 Suzhou

36

Summary

- Metrics and evaluation should be User-centred
- Combining qualitative and quantitative measures
- Typical scenarios are very important for the community.
- Multi-disciplinarity in the evaluation but also the development.