Benchmarking Usability of Assistive Robotic Systems: Methodology and Application

Juan G. Victores, Alberto Jardón, Fabio Bonsignorio, Martin F. Stoelen, and Carlos Balaguer

Abstract—**Usability can be defined as the degree of a product's fitting to the characteristics of a person or of a group of people. The concept of usability includes aspects of using a product that are closely linked to the user's degree of satisfaction and preferences. As a multidisciplinary concept, definitions may vary depending upon the specific area on which one focuses. However, common terms can be found throughout literature. Parameters such as the difficulty and steepness of the learning curve for the end-user, or flexibility and adaptability are commonly evaluated. In the context of Assistive Robots, factors taken into account are related to user acceptance, security, precision of task execution, and overall system efficiency. Hence, it is also closely linked to the concept of dependability. Boundary conditions related to the environment and to the user must be taken into account. In this paper, the importance of the role of benchmarking the usability of Assistive Robots is discussed, and a methodology for obtaining usability data from experiments is proposed. The proposed methodology is part of a continuous improvement framework that is based on the System Knowledge Space, which will be described within the text. Then, a general view at results extracted from experiments performed with an Assistive Robot and real potential system end-users in realistic scenarios is given. This exemplary usability benchmarking assessment follows the guidelines of the methodology that is proposed. The experiments that are described were developed as part of the ASIBOT program at the Universidad Carlos III de Madrid in collaboration with the National Paraplegic Hospital in Toledo (Hospital Nacional de Parapléjicos de Toledo). The last part of our paper deals with results of how these experiences have influenced actual and future research efforts and discusses how this should positively affect the scientific research and developer community.**

Index Terms—**Benchmarking, Clinical Trials, System Knowledge, Usability.**

I. Introduction

THE current technological boom and the population's generally high expectations of technology favor the creation and acceptance of new products. More is expected of new technological developments within the field of disability than by society at large. Users believe and expect they will be able to overcome their disability by using alternatives arising from the development of technical aids designed for and adapted to their needs. In the disability sector, there is a steady stream of new products that interact with the user in a higher or lower degree in order to meet their everyday needs. Like any product used regularly, technical aids must satisfy user's expectations regarding the functions they hope to perform. In the case of products designed for the field of disabled persons, this requirement becomes even more important, recalling that the functions replaced are generally essential for personal autonomy.

There are numerous testing procedures and/or evaluation standards that measure how well a product performs its functions in terms of technical quality regulations, which can evaluate properties such as hardness, durability, certain kinds of safety (electrical or thermal isolation), or other technical evaluation parameters more specific to robot technology, such as repeatability or manipulability. These standards and procedures are related to the characteristics of the technical aid itself or, at most, to the user's relationship with the product. These considerations are useful, but only part of an initial design stage, and remain static and unchanged independently of the user's needs or preferences.

User satisfaction concerning technology, in addition to fulfilling their expectations of technological resources, is conditioned by the emotional perception those resources provoke on users. This aspect is frequently forgotten when designing and implementing a resource. Although the market is starting to realize how very important it is to incorporate strategies that include the final user, these techniques are not always applied, or at least the way they are carried out is of questionable value [1][2]. Efforts related to the usability of user interfaces are being currently being performed by standardization organisms and special interest groups [3][4]. However, these standards and groups are highly related to software and interfaces, much in the terrain of pure accessibility. Examples of usability studies in robotics can be found in the domain of rehabilitation robotics. Simpson *et al.* were able to identify the limitations and ways to improve their assistive wheelchair navigation system [5]. Usability assessments for manipulator interfaces with patients with physical disabilities were performed by Buler *et al.*, Parsons *et al*. and Keates *et al* [6][7][8]. However, a lack of methodology for defining their experiments has been detected, leaving doubts in the validity of their results [9].

Manuscript received February 15, 2010.

J. G. Victores (corresponding author), A. Jardón, M. F. Stoelen, and C. Balaguer are members of the RoboticsLab research group within the Department of Systems and Automation, Universidad Carlos III de Madrid, (jcgvicto, ajardon, mstoelen, Balaguer)@ing.uc3m.es

Fabio Bonsignorio is a member of the RoboticsLab as Santander Chair of Excellence in Robotics at Universidad Carlos III de Madrid and is CEO/founder of Heron Robots of Genoa, Italy, fabio.bonsignorio@uc3m.es

This paper discusses the role and importance of benchmarking usability in the context of our own framework for continuous improvement of the state of the art of Assistive Robots. This framework may not be the best, and it is at most not completely mature, but is set as a starting point to structure our desires and needs to fulfill common goals in providing efficient, usable and dependable systems. A developed methodology proposal is described in the context of the framework. Finally, we present our own experience at assessing the usability of ASIBOT, a prototype of a personal assistance robot, involving users directly, as specified by the methodology. This study is part of ASIBOT, a new robotic technology development project to benefit the disabled. Our research was carried out at the National Paraplegic Hospital in Toledo, a center that specializes in comprehensive treatment for persons with spinal cord injuries, a physical affection that affects the mind's mechanisms of voluntary movement generation, leading to severe disability. The study focused on planning the experiment with patients associated with this Center. The ASIBOT project for developing and experimenting with a portable personal robot to aid disabled and elderly people with everyday activities was approved in 2005 by IMSERSO as part of its scientific development, research and innovation in support technology for disabled and elderly people.

II. Continuous Improvement Framework

Our current framework for continuous improvement is based on the System Knowledge Space (SKS). A representation of the SKS can be seen on Fig. 1. It is the space of all the possible knowledge of a system, independently from who possesses that knowledge. The objective of the framework in this field is to devise methodologies and design patterns to transfer knowledge into the realm of the scientific developer community. Current research efforts are focused on:

a) Methodologies for transferring End-User Knowledge into the Developer Knowledge subspace and quantifying the results. This may usually be accomplished through successive phases of experiments and feedback.

b) Design patterns for implementing successful interfaces or algorithms, for similar objectives.

A more precise understanding of the possible underlying mechanisms that can be invoked when applying our continuous improvement framework may be achieved throughout the description of the proposed methodology.

III. Proposed Methodology

The proposed methodology for benchmarking usability is composed by three phases. Phases II and III are each divided into two steps, according to the different types of transference of knowledge that should occur in the SKS. There is an initial phase in which no transference of knowledge in the SKS is intended.

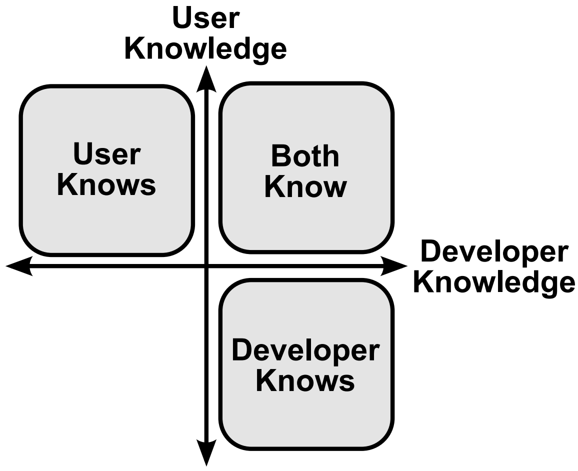

Fig. 1. Four subspaces of the System Knowledge Space. The intention of the diagram is to divide the space into four subspaces, where all possible elements of the space are contained, such as in set theory.

A diagram of the Phases and Steps that compose the proposed methodology for benchmarking usability of Assistive Robots is depicted on Fig. 2.

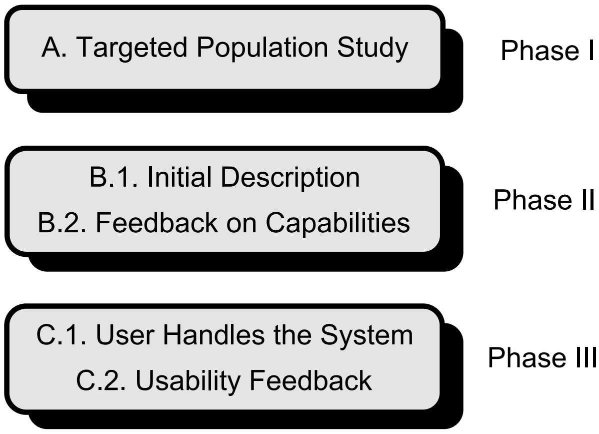

Fig. 2. The three Phases of the proposed methodology for benchmarking usability of Assistive Robots.

A. Phase I: Targeted Population Study

The description of user inclusion and exclusion criteria should be clear, concise, and explicit. If the reason for selection criteria should be considered not evident, it should be justified, either by means of an extended explanation, or reference to previous scientific publications. Implicit in the selection of a targeted population study is the need for members inside the developer group with a certain degree of expertise in the matter of the subjects with whom they will be treating. This may call for the need of specialized psychiatrists when treating with users with mental disorders, or perhaps pediatricians if treating with infants, children or adolescents. A common language is also a necessary element that is many times overlooked.

Although this Phase does not imply transfer of knowledge in the SKS, it is worthy of mention, as it is a prerequisite for the performance of the successive Phases of the methodology. Poor decisions in this Phase can later result in difficulty or inefficiency of the performed experiments.

B. *Phase II: Initial Description, Feedback on Capabilities*

In this Phase, the system should be explained to the end-users, and feedback received. It is composed two Steps.

B.1. Phase II, Step 1: Initial Description

In this Step, the flow of knowledge from the "Developer Knows" subspace to the "Both Know" subspace of the SKS should occur. An initial description of the system and system components is given to the targeted population. Explained elements of the system should include the possible interfaces to the system, and a global view of the details of interest of the actual Assistive Robot. Media included for this presentation may be slide-show presentations, videos, or actual demonstrations of the real working robot. Various paradigm approaches can be used for this phase, where most important variables are time and distance. Demonstrations may be 'in situ', which means same time and place, or may occur in different time or distance slots using modern technologies such as on-line video streams of pre-recorded presentations or video-conference.

B.2. Phase II, Step 2: Feedback on Capabilities

This Step is where the transition of knowledge from the "User Knows" subspace to the "Both Know" subspace in the SKS occurs. It is important to notice how the activity of the previous step allows this otherwise impossible flow of information (unless the system is known from the media, etc., which would mean there was a previous flow of knowledge analogous to the last Step). Here, end-users are given questionnaires designed by the developers paying special attention to the recommendations set by the experts on the target population group. The scope of the questionnaires is to gather information about the possibilities the end-user can devise in the product. The scope should be open, allowing the user to express current need and also futuristic desires. However, the total length of the test should not be too extensive, to avoid stress and boredom. Questionnaires, especially in the case of users at a distance, may be performed on-line. The process may be optimized by the automation of the transference of results into a centralized data-base. This data-base should be shared by the community of developers and serve for comparing and learning from each other's results, a fundamental part of benchmarking.

C. *Phase III: User Handles the System, Usability Feedback*

From the previous Phase, we have information of the ways the end-users would like to interact with the system, and have generated a list of preferred user tasks. In Phase III, the scenario is set to fit with the users' desires (as much as technically possible) and interact with the system. After the trial, they provide feedback. The following is a more detailed explanation of the two Steps that compose the Phase.

C.1. Phase III, Step 1: User Handles the System

Here, knowledge flows from the unknown subspace to the "User Knows" subspace in the SKS. It is vital to understand that this one of the fundamental mechanisms to obtain the final benchmarks of usability.

In this Step, the paradigm place of time or distance shift may also be applied, but its implementation may be less evident or more difficult. If long-distance tele-operation is unavailable or undesired, emulated interfaces and simulations of robot behavior in immersive 3D scenarios may be necessary. The modes of operation experimented by the user should be as abundant as possible. This way, a higher range of information about system usability will be received. However, it is important to vary only one parameter at a time, and clearly tag the information related to each experiment. Factors such as sex, skill, age, or degree or kind of disability may create systematic bias and interfere with the final results.

C.2. Phase III, Step 2: Usability Feedback

This is the Step of flow of the Usability knowledge. Knowledge is transferred from the "User Knows" subspace to the "Both Know" subspace. For this purpose, questionnaires similar to the ones described for the Feedback on Capabilities Step should be handed to be filled by users. The premises to be taken are similar. Here, on-line questionnaires are highly recommended, as they can easily be adapted to interface with a global knowledge of Usability data-base. Again, this data-base should be publically available through permanent links for the benefit of the community. While this kind of data-base is temporarily unavailable or non-existent, it is convenient to publish results in an orderly fashion as articles on relevant journal or conference publications.

A summary diagram of the flow of knowledge in the SKS throughout the different Phases and Steps of the proposed methodology can be seen in Fig. 3.

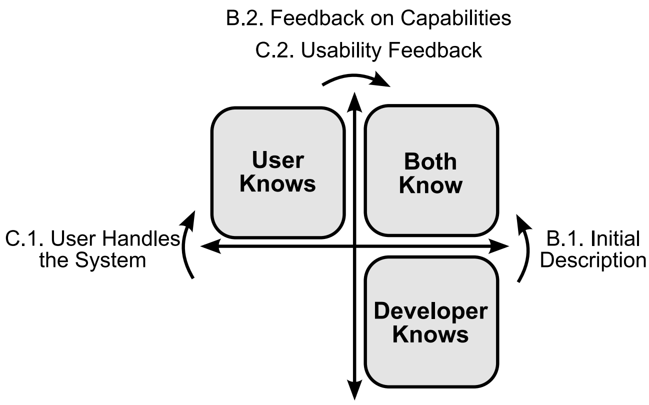

Fig. 3. Flow of knowledge that occur throughout the Phases of the proposed methodology for benchmarking usability.

IV. Experiment (Methodology Application)

In this section, an application of the proposed methodology for Usability Benchmarking will be explained. It is an elaboration on gathering information from the interactions of selected users with the robot in a structured format, as implicit in the proposal.

Details of our research procedure were planned by reviewing the relevant bibliography and evaluating several tests related to developing a methodology for assessing user satisfaction with devices for assisting disabled people [10][11]. In our approach to the design, we incorporated the concept of usability. We decided that the technique or tool used to evaluate the product should provide information about individual representative users via consecutive interventions by several of them. These users would benefit directly from the assistance and information would be gathered in their presence [12][13]. In the end, out of all the techniques we reviewed that permit user participation and given the lack of a system to evaluate user satisfaction that was appropriate for our needs, we decided to base our design on the following techniques: “Usability tests” supported by “Creating settings” and the application of evaluation criteria as defined in what is known as “The K.I.U. Test” [14]. The main aims of these techniques refer to identifying frequent major problems; detecting errors, needs, or requirements; generating design criteria and final user requirements; as well as a global usability assessment (detecting lacks, possible causes, and proposing solutions) of the product, which allows us to identify which aspects need to be modified in the new design. The main deficiencies in usability refer to criteria of ease of learning to use it, utility, functionality, ease of use, and user satisfaction [2][15][16].

A. Phase I: Targeted Population Study

The target population studied was people who had had spinal cord injuries for at least a year. No cases with acute injuries were considered. We looked for users who, once past the initial phase, had spent regular periods of time in their homes, which gave them a perception of the main difficulties to be found in their daily lives. Because of their daily experience in facing numerous problems of dependency, they were able to evaluate the functionality of technical aids with more objectivity.

The robot was designed to assist severely disabled people. For that reason, people affected at the cervical level were chosen, from C4 to C8 neurological levels, because of the resulting limitations in their upper extremities and the chance of doing the different tests without problems of vital capacity. Exclusion criteria were: epilepsy, mental retardation, uncorrected visual deficiency, or psychiatric problems. The group analyzed in the final tests consisted of six users who fulfilled the above criteria.

B.1. Phase II, Step 1: Initial Description

A general information hand-out about ASIBOT was distributed among the attendants, and a multimedia presentation session was held (see Fig. 4). After the presentation, the users could ask questions to the developer team to clarify details they considered necessary about the system. Since the presentation had been simultaneously streamed on-line, users who had not been able to physically attend were able to ask questions thanks to a tele-conference mechanism. This assured the group of users could obtain sufficient knowledge about this climbing and portable Assistive Robot, ASIBOT.

Fig. 4. Initial presentation to users about ASIBOT.

B.1. Phase 2, Step 2: Feedback on Capabilities

After gathering data in the first face-to-face session with six users who met the inclusion criteria, information was elicited via questions on a questionnaire, in order to focus subsequent real tests with the robot on the most frequently demanded activities. In one of the first questions about their main demands for independence, the users were able to state which activities they find most unpleasant and would like to be able to do without depending on another person, regardless of whether the robot could do them. Ranked in order of importance, lists were compiled of the main activities the user finds most unpleasant where the user depends on another person. Users were also asked how useful they found the presentations and videos, for future improvement.

B.1. Phase III, Step 1: User Handles the System

The tests were held in a hospital context, carried out in a setting appropriate for carrying out the activities to be studied. Facilities at the hospital’s Occupational Therapy Unit were used that partially simulated an everyday home environment. We tried to create settings for the actual tests with the robot that met users’ demands as close as possible and that were within the robot’s performance capabilities. Subsequently, depending on each setting, the most appropriate of the available functions were chosen for the tests. The results from the previous Phase showed that personal hygiene was the most appropriate for the tests carried out with the robot. That was why an assessment was carried out in a bathroom setting. The assessment of available resources for the tests also led to our decision to use the bathroom facing the mirror, given the chances it offered to assess a larger number of basic functions in everyday life. We proposed four functions in this setting: drinking, putting make-up on, brushing one’s teeth, and washing one’s face. Fig. 5 depicts the Bathroom Setting created.

The work team made up of staff from both participating entities thought it best to carry out a pilot test prior to the tests in real settings in order to assess different ways of interacting with the robot-user interface implemented in a PDA where, in the shape of large visual-tactile buttons, six different options (four related to the bathroom setting) were given for the response or function the robot was to perform. The following

Fig. 5. User commanding the robot.

choice modes were offered ranked from the largest to the smallest mobility requirements to activate them:

- Tactile, via the user's touch or a pencil.
- Joystick, to choose the option with a button to validate.
- Voice recognition, different options for activation and choice by voice.
- Lighted sequence, where the options are highlighted one at a time. User selection is performed by pressing a single chin selection button, joystick top button, or commanding by voice when the desired option is lit.

B.1. Phase III, Step 2: Usability Feedback

We incorporated the techniques referred to at the beginning of this section into the design of our own questionnaire, made by the biomechanics and technical aids research and development staff at the National Paraplegic Hospital in Toledo. In a pilot study, the validity of the data and understanding of the questions were verified. The questions were analyzed to be certain they were appropriate for our objectives and we verified that the total length of the test was not more than 30 minutes, in order to prevent fatigue or distraction. Questions that evoked negative stimuli in the user were eliminated and explanations on how to fill in the questionnaire were included. The user was instructed to answer quickly, without spending much time thinking about each answer. The questionnaires were anonymous. If needed, the health care staff at the Biomechanics Unit helped to fill it in. No personal data were gathered which identified any patient except for the level of spinal cord injury. This fact might be of interest in identifying the patient's residual capacities and relating them to the answers. The questionnaire for the tests had three parts: A first part with closed questions, with answers given on a Likert summative five point satisfaction scale (from +2 to -2) and focused on an assessment of the functions or activities suggested for the robot to perform during the tests. The second part consisted of questions in the former format focused on an assessment of the robot's characteristics. In the third part, four open questions were asked about its use, utility, and needs for assistance of each individual. Lastly, a general assessment of technical aid was made through two specific questions with answers in an ordinal scale format.

V. Results from Experiment

This section is a survey on most important results extracted from the experiment. The following is a description of the feedback received from each feedback step of the methodology.

A. Feedback from Phase II, Step 2: Possibility Feedback

The majority of users demanded getting dressed and washing themselves, in that order. The results in order of importance when asked to prioritize four settings or situations that we proposed where using the robot would be most useful were the following.

1. Daily hygiene (washing their face and hands, brushing their teeth, combing their hair, shaving, or applying make-up).
2. Lying in bed (bringing objects near...)
3. In the wheelchair (eating, drinking, bringing objects near...)
4. In the kitchen (opening cupboard doors, moving utensils...)

After evaluating the different interfaces offered, the following conclusions were drawn from the six users' answers:

- Except for one person, they were all seen to be capable of HANDLING ON THEIR OWN the different interfaces offered: voice recognition (with different options), joystick, tactile, and lighted sequence with a selection button.
- Regarding EASE OF USE, they chose in order of preference: tactile, voice recognition, joystick, and lighted sequence.
- The most PRACTICAL interface turned out to be tactile, followed by voice and tactile in the same number of responses, and lastly, the lighted sequence.
- The interfaces based on VOICE recognition and the JOYSTICK were preferred regarding APPROPRIATENESS of the interface given their MOVEMENT ABILITIES.

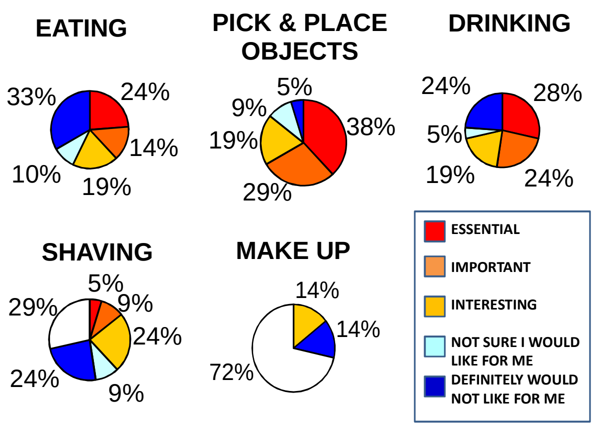

Fig. 6. User Feedback regarding Possibilities of Use.

B. Feedback from Phase 3, Step 2: Usability Feedback

Four of the six users from the previous day were present, plus one new user who met all the inclusion criteria, although he wasn't there for Phase II. They were all male. The functions of drinking, brushing their teeth, and washing their faces were assessed. For each of the functions, they were instructed to assign a response score to a set of statements on a Likert satisfaction scale from –2 to +2, where +2 is "completely satisfied", +1 is "satisfied", 0 is "not sure", -1 is "dissatisfied", and -2 is "completely dissatisfied".

The global results for each of the three functions were as follows:

- In the function of **DRINKING**, the average score was **0.75** on the Likert scale. One of the users did not answer for this function.

- In the function of **BRUSHING ONE'S TEETH**, the average score was **0.46**.

- In the function of **WASHING ONE'S FACE**, the average score was **-0.86**.

CONCLUSIONS

Benchmarking is not only a "hot topic", it is extremely necessary. Usability, on the other hand, is an extremely necessary benchmark for Assistive Robots. This paper proposes a methodology for benchmarking Usability in the context of a continuous improvement framework. This methodology can be used as a design pattern that can be merged with other important parameters in Assistive Robotic Systems such as timing, number of user events for task achievement, tolerance to dimensional errors, tries/errors ratio, or steepness of learning curve. However, as established from the beginning of this text, we believe that the user's opinion counts and is actually an empirical data source that should be formally collected and shared throughout the community. We are optimistically hoping the union of sufficient critical mass to devise portals and data-bases with collective usability benchmarks in an open-source project way, well parameterized and permanently linked. CAD models of the disposition of elements used during experiments could be uploaded. The exposed example case is based on a climbing robot. However, we foresee soon visioning comparisons of our statistics with those of researchers working with relevantly different types of robots, all in the field of Assistive Robotics.

ACKNOWLEDGMENT

The authors would like to thank all the ASIBOT team at Universidad Carlos III de Madrid, and the fellow collaborators at the Hospital Nacional de Parapléjicos de Toledo.

REFERENCES

[1] C. Vicent and L. Demers, "Electric beds: Recommendations based on Users' Satisfaction," *Occupational Therapy*, March, 2002.
[2] M.J. Fuhrer, "Assistive technology outcomes research: Challenges met and yet unmet," *American Journal of Physical Medicine & Rehabilitation*, vol. 80, 2001, pp. 528-535.
[3] *Guidance on Usability*, ISO Standard 9241-11, 1998.
[4] http://www.stcsig.org/usability, retrieved February 2010.
[5] R. Simpson, D. Poirot, and M.F. Baxter, "Evaluation of the Hephaestus smart wheelchair system," *Proc. of the 6th Int. Conf. on Rehabilitation Robotics*, Stanford, CT, 1999, pp. 99-105.
[6] C. Bühler, R. Hoelper, H. Hoyer, and W. Humann, "Autonomous robot technology for advanced wheelchair and robotic aids for people with disabilities," *Robotics and autonomous systems*, vol. 14, 1995, pp. 213-222.
[7] B. Parsons, A. White, P. Warner, and R. Gill, "Validation methods for an accessible user interface for a rehabilitation robot," *Universal Access in the Information Society*, vol. 5, 2006, pp. 306-324.
[8] S. Keates, J. Clarkson, and P. Robinson, "Designing a usable interface for an interactive robot," *Proc. 6th Int. Conf. Rehabilitation Robotics*, pp. 156-162.
[9] Y. Endo, D.C. MacKenzie, and R.C. Arkin, "Usability evaluation of high-level user assistance for robot mission specification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, 2004, pp. 168-180.
[10] J.W. Jutai, M.J. Fuhrer, L. Demers, M.J. Scherer, and F. DeRuyter, "Toward a taxonomy of assistive technology device outcomes," *American Journal of Physical Medicine & Rehabilitation*, vol. 84, 2005, pp. 294-302.
[11] L. Demers, R. Weiss-Lambrou, and B. Ska, "The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress," *Technology and Disability*, vol. 14, 2002, pp. 101-105.
[12] M.A. López, R. Poveda, I. Andreu, R. Barberá, P. Ramiro et all. Datus. *Guía Práctica para fabricantes de productos de la vida diaria y ayudas técnicas*. CEDAT and IBV, 2003.
[13] L. Demers, R.D. Wessels, R. Weiss-Lambrou, B. Ska, and L.P. De Witte, "An international content validation of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST)," *Occupational Therapy International*, vol. 6, 1999, pp. 159-175.
[14] Instituto de Biomecánica de Valencia, *Método para la selección de ayudas técnicas bajo criterios de usabilidad*. Musa/IBV, 2003.
[15] L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, and J.P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *NATURE-LONDON*, vol. 442, 2006, pp. 164-171.
[16] A. Granados, "Herramientas y Metodología en la Evaluación de Tecnologías Sanitarias," *NETS, Suplemento*, vol. 1, 1998, pp. 33-38.