SRS

Multi-Role Shadow Robotic System for Independent Living

Small or medium scale focused research project (STREP)

DELIVERABLE D6.2

SRS user validation - results

Contract number :	247772
Project acronym :	SRS
Project title :	Multi-Role Shadow Robotic System for Independent Living

Deliverable number :	**D6.2**
Nature :	R – Report
Dissemination level :	PU – PUBLIC
Delivery date :	18-04-2013

Author(s) :	Lucia Pigini, Marcus Mast, David Facal, Alexander Noyvirt, Renxi Qiu, Claudia Salatino, Alvaro Garcia, Rafael Lopez
Partners contributed :	INGEMA, FDCGO, HdM, HP, IPA, ROBOTNIK, CU
Contact :	csalatino@dongnocchi.it

The SRS project was funded by the European Commission under the 7th Framework Programme (FP7) – Challenges 7: Independent living, inclusion and Governance

Document History:

Version	Author(s)	Date	Changes
ToC	Lucia Pigini	13/02/2012	
V1	Lucia Pigini, Marcus Mast, David Facal, Alexander Noyvirt, Davide Sommacampagna	20/06/2012	First version
V2	Lucia Pigini,	30/06/2012	Added in chapter 4.3
V3	Davide Sommacampagna	30/06/2012	Additional in chapter 4
V4	Renxi Qiu	01/07/2012	Merge and quality Check
V5	Lucia Pigini	02/07/20012	Interim version ready for submission
V6	Alvaro Garcia	09/04/2013	Added in chapter 5
V7	Alvaro Garcia, Claudia Salatino	15/04/2013	Chapter 5 revision
V8	Marcus Mast	17/04/2013	Added in chapter 6
V9	Claudia Salatino	23/04/2013	Added in Robotnik contribution, conclusions; final revision
Final Version	Renxi Qiu	17-05-2013	Ready for release

EXECUTIVE SUMMARY

The aim of the evaluation phase with potential users is to investigate and measure effectiveness, usability and acceptability of the advanced prototype to generate feedback for improvement.

This document reports results of the validation process of the SRS prototype with users.
It reports results achieved so far in five main SRS testing steps involving users:

- the first one specifically concentrating on arm manipulation and UI-PRO visualization, led by INGEMA and ROBOTNIK and carried out in San Sebastian;

- the second one, carried out in Stuttgart an led by HDM and IPA concentrating on the technical issues emerging in a real home environment, a necessary pilot study to prepare the main set of tests of the whole prototype;

- the third one, carried out in Milan and led by FDGCO and CU. It collected full information of the integrated prototype related both to potential users perception and to technical effectiveness. Results showed high level of interest with respect to the potential of the system and high level of acceptance. Nevertheless the results of Milan tests showed that the state of SRS system was not fully sufficient to make people consider it as a product ready for adoption, and also usability of interfaces could be further improved especially on UI-PRO. The interface obtained lower scores on successful indicators specified in D6.1. For the reason, an further investment of technical development was carried out and two additional iterations of testing were planned as follows;

- the fourth one, carried out in San Sebastian by INGEMA and ROBOTNIK, specifically concentrating on arm manipulation and UI-PRO visualization with the interface updated after the first trials. During San Sebastian tests, objective measurements - regarding task completion, number of errors made, number of questions asked by users, hints needed - showed the improvement of the system in the usability related measurements, with one exception on time. Mainly, the slow speed of tasks' performance can be frustrating for end users. With SRS development, the number of functionalities of the robot was significant expanded, the robot can currently automatically detect the object, gets near to the object in an optimal and safe position, etc. All of these processes have their computational cost that reflected in the execution time;

- the final one, carried out in Stuttgart by FRAUNHOFER, HDM, BUT, concentrating on the further UI-PRO developed, specifically on the weakness identified in the Milan test. During Stuttgart tests users were able to complete manipulation and navigation tasks in 100% of the cases, meaning that the user interface was highly effective. Results on the system's usefulness are encouraging and show a vast improvement over the previous study. All average ratings were well above neutral (4) with some reaching close to maximum values. With the final test, successful indicators specified in SRS D6.1 has been achieved with the project period.

Overall the SRS system was perceived interesting and potentially very useful by all user groups during the succession of tests performed. Acceptance is high and this can be highlighted as the result obtained in tests with the integrated prototype. Elderly people and private caregivers were more enthusiastic about the novelty represented by SRS system; professional operators were also impressed with the latest revision of the SRS interfaces using Space Navigator and Stereo Vision.

TABLE OF CONTENT

1 INTRODUCTORY SECTION: PURPOSE OF DOCUMENT AND CONTENTS

The present document reports a detailed description of user validation results achieved. The validation phase was designed in a way that allowed investigating effectiveness, usability, and acceptability during incremental and complementary stages by analyzing some specific features and the entire prototype, so as to generate feedback for improvement.

The test sites settings predisposition, plan for ethical and safety issues, research questions addressed, experimental protocol description, and validation methods adopted were already reported in D6.1-b.

Each chapter of the present document reports a detailed description of results obtained in each test session; in particular:

Chapter 2 reports arm manipulation and RO-PRO visualization tests' results, carried out in San Sebastian, to address the main behavior of the robot with real users.

Chapter 3 reports the first real home environment case test results, carried out in Stuttgart, to address first SRS experience outside the laboratory and the first users impressions.

Chapter 4 reports results about advanced SRS prototype tests, carried out in Milan, to address the integrated SRS functionalities and the scenarios effectiveness with a sample of real users in a home environment.

Chapter 5 reports results of the updated UI-PRO interface tests, carried out in San Sebastian by INGEMA and ROBOTNIK, specifically concentrated on arm manipulation and UI-PRO visualization. After the first trials (chapt. 2), several usability suggestions were made. The interface was adapted according to the suggestions in close contact between user partners and technical partners.

Chapter 6 reports results of the final test session, carried out at Fraunhofer IPA in Stuttgart, concentrating on UI-PRO developed as part of the EEU extension to obtain user opinions and ratings, determine usability issues, and assess the overall utility and suitability of the solution for remote navigation and manipulation by operators in a call center.

Chapter 7 reports conclusions of the deliverable.

2 MANIPULATION TESTS AND VISUALIZATION TESTS

2.1 MANIPULATION TRIALS

Manipulation and visualization tests in San Sebastian were conducted in January 2012 with the aim of address the main peculiarities of the robotic arm with real users.

For the trials, according to mobility problems of the participants, the Rehabilitation Service facilities were prepared for on-site testing. In concrete, we used one of the extensive exercise rehabilitation gymnasiums of the Matia Hospital (Figure 1). The rehabilitation gymnasium is oriented at the physiotherapy and rehabilitation of patients of early injured elderly people. This service aims to provide support and professional training to patients that have underwent stroke, ictus, hip fracture, lower and upper limbs fracture, traumatic injuries, etc.

FIG. 1 MATIA HOSPITAL'S REHABILITATION GYMNASIUM

In these trials, we have combined ad-hoc questions with a validated test, the AttrakDiff (Laugwitz, Held, & Schrepp, 2008), selected to measure user experience in a simple and immediate manner (details about protocol development can be consulted in D6.1).

2.1.1 PARTICIPANTS

In the Manipulation trials, we collected data about frail elderly users' subjective perception on the robotic arm on tasks based on grasping movements. We recruited 14 patients of the Rehabilitation Service of Birmingham Hospital - Matia Foundation. They were 10 woman and 4 men. Their mean age was 79.14 (st. dev. = 9.607), maximum 92 years and minimum 60 years . They studied a mean of 8.79 years (st. dev. = 4.21), maximum 17 years and minimum 4 years. As expected, men were slightly younger than women (Men 71.75 years, st. dev. = 12.53; women 82.10 years, st. dev. = 6.84), although years of education were similar (Men8,23 years, st. dev. = 6.13; women 9 years, st. dev. = 3.59).

All of them presented mobility problems in lower limbs, and some of them also in upper limbs. As reported in D6.1, we have used two frailty index in order to compare them: the Barber test, used previously in other user studies within this project, and the SHARE-FI, a more recent and potentially more accurate index based on data from SHARE - Survey of Health, Ageing and Retirement in Europe. Mean score in the Barber test was 3.36 (st. dev. = 1.01, maximum 5, minimum 1). Apart for the maximum score in hospitalization, 42,9% of the participants live on their own, 21,4 % do not have a relative to call on for help, 50% depend on someone for regular help, 64,3% are concerned about their health, 28,6% have difficulties with vision and the same percentage have difficulties with hearing. None of the participants reported problems to have a hot meal and only one participant, the one with the higher frailty score, reported to be confined to her home.

Mean score in the SHARE Frailty Index was 2,51, with 7 participants obtaining a Frail score, 5 participants a Pre-Frail score and 2 Non-Frail. Nevertheless, the SHARE-FI index presents some limitations in this population, especially because of the high weight of the item about frequency of physical exercise, which is high in persons attending a rehabilitation service. In line with this observation, SHARE-FI score correlates only moderately with the total score in the Barber test (r=0.352).

Some participants could present mild cognitive impairments, although only one presented suspicious of possible dementia (Mini-Mental State Examination < 23). Regarding their subjective perceptions, they provided average responses about their memory capacity compared to the rest of the society and with their capacity in the past.

As expected, most of them use TV and telephone / mobile phone daily; they use washing machines, dishwashers and vacuum cleaners frequently; and they never or very rarely use smart phones, PC or laptops.

2.1.2 Usability and accessibility

Regarding the usability and accessibility assessment, we should take into account that these trials only shown the capabilities of the system to the users instead letting them use it by themselves. This demonstration of the system's usability and accessibility becomes directly related with the acceptability. However we will distinguish this part related to the direct usability scenario acceptability and will develop below the acceptability concept as the general system's acceptability concept.
The system was presented to the user catching a cocoa storage jar in front of him (Figure 2). The system worked well in 12 out of 14 times catching the tin. 2 times the system needed to be restarted because it was jammed grabbing the tin; and worked well 13 out of 14 times releasing the tin. The system dropped the tin in the remaining one.

FIG. 2 MANIPULATION TRIALS AT MATIA HOSPITAL'S REHABILITATION GYMNASIUM

One observational category with three conditions was established to define the users' emotional reaction to the system. Verbal, facial expression and position were used to establish the emotional reaction. Relaxed, Nervous and Scared were the three observational categories chosen. The users were all relaxed in front of the robotic arm. Apart from that, several users reacted to the arm by trying to touch the fingers or catch the tin, showing curiosity for the system.

When the users were confronted to the robot arm functionality several users stated their opinions spontaneously, the others that did not stated, were asked specifically. Below, we can find the users verbal behavior:

- *"It's very interesting".*
- *"It should bring the stuff nearer".* He is confident that the system works well. *"The ideal would be to have this thing that brings everything nearer. It could feed me, but that would need more care and support. It should have wheels"*
- *"It is not bad for a person that lives alone. With a child at home it would be as a toy"*
- *"In my situation is useless. We live three brothers at home, one with Alzheimer's disease, one widow and me. All of we older than 80 years old. In our home it would be useless. Fantastic for others, a good invention that can be very useful (for others)".*
- The user stated that do not see the system useful, at last meanwhile he was able to perform the activities autonomously.
- *"What a smart thing! There are so many good inventions today…"*
- The user commented the movement while the robotic arm was performing. The robotic arm seems useful to him. He states that he would use it but only if he has money to buy it.
- *"Everything modern is good. It would support health"*
- *"It would be useful for several things"* He would prefer to interact directly with the system instead of having other people controlling it.
- *"Meanwhile there was a person that is not able to do things by himself he likes it. The following step is to get back the stuff to its place. It is useful for a person that lives alone".*
- *"Cool".* The user liked it and asked questions about the robotic arm.
- *"It's marvelous, but it cannot be at home. It seems to be a toy and it's funny to see how it works."* She does not think that the system is useful. She stated that to have a deeper understanding of the system she should see the system working in more situations.
- *"The system is fantastic but gigantic".* The user stated that the system would be very useful for him, and ask more of the system capabilities.

We can concretize qualitative information collected in the following topics:

- The system is perceived as useful, as far as the person lives alone and is not fully autonomous. Although all the participants have daily living difficulties related to their mobility problems, they still claim that the system will be only useful *"meanwhile there are a person that is not able to do things by himself"* (or herself), showing a certain perception of autonomy even in dependent older adults.
- The system is interesting, interest on technology.
- Participants point to some difficulties (i.e. the size of the robot and its cost).

2.1.3 Acceptance

In spite of the peculiar appearance and technological restrictions of the robotic arm presented, the system did not evoke responses of fear or rejection in the participants (see Table 1). Users do not seem afraid of the system and their behavior shows that they think that the system is safe.

TAB. 1 FREQUENCY OF FRAIL OLDER ADULTS' RESPONSES IN THE ATTRAKDIFF FOR THE MANIPULATION TEST.

	1	2	3	4	5	6	7	
Annoying	-	-	-	4	1	7	2	**Enjoyable**
not understandable	7	4	-	3	-	-	-	**Understandable**
creative	8	5	1	-	-	-	-	**Dull**
easy to learn	5	5	1	-	1	2	-	**difficult to learn**
valuable	9	1	2	1	-	1	-	**inferior**
boring	1	-	-	2	4	4	3	**exciting**
not interesting	1	-	-	1		5	7	**interesting**
unpredictable	-	2	-	-	3	3	5	**predictable**
Fast	4	3	1	5	1	-	-	**slow**
inventive	4	6	1	2	1	-	-	**conventional**
obstructive	-	-	-	4	1	2	7	**supportive**
good	7	4	1	1	-	1	-	**bad**
complicated	-	1	1	2	4	4	2	**easy**
unlikable	-	1	-	5	1	5	2	**pleasing**
usual	-	1	-	1	2	5	5	**leading edge**
pleasant	3	5	-	5	-	1	-	**unpleasant**
secure	6	6	-	2	-	-	-	**not secure**
motivating	3	5	1	2	1	-	1	**demotivating**
meets expectations	5	6	1	1	-	-		**does not meet expects.**
efficient	7	4	3	-	-	-	-	**inefficient**
Clear	6	4	3	-	-	-	-	**confusing**
impractical	-	-	-	2	1	5	5	**practical**
organized	4	6	-	2	-	1	-	**cluttered**
attractive	4	8	-	-	5	-	-	**unattractive**
Friendly	2	4	2	3	1	-	-	**unfriendly**
conservative	-	1	-	2	-	7	3	**innovative**

Regarding acceptance ad-hoc questions (Table 2), we obtained:

- A higher percentage of participants with positive responses (Likely and Very likely) in items 1, 2, 3 and 6, showing a majority of positive perceptions about the functionalities of the robotic arm and the intention to use it.

- A higher percentage of participants with negative responses (Very unlikely and Unlikely) in items 4 and 6. Negative responses to item 4 ("Not have to do everything by myself") could reflect the importance of autonomy and sense of independence in this population. Negative responses to item 5 ("I like to use this kind of appliances") seem to points to a lack of relation between acceptance and subjective experience with technology, that it is going to be analyzed in more detail in D6.2.

- An equal percentage of participants with positive and negative responses to item 7 ("I would buy this system when it becomes available"). This result should be taken into account in task 6.4 SRS Cost-effectiveness Assessment & Socio-economic Implications.

TAB. 2 FREQUENCY OF FRAIL OLDER ADULTS' RESPONSES TO ACCEPTABILITY ISSUES IN THE MANIPULATION TEST

	Very unlikely	Unlikely	Neutral	Likely	Very likely
1. In general, would you use the functionalities?	3	1	1	5	3
2. The robotic arm can help me to achieve goals in my daily routine	2	-	3	4	5
3. More control over my daily life.	4	-	4		6
4. Not have to do everything by myself	5	2	3	3	1
5. I like to use these kind of appliances	5	2	4	3	-
6. I would use this system when it becomes available	2	1	4	6	1
7. I would buy this system when it becomes available	5	1	2	3	3

Regarding qualitative information, several users stated opinions expanding their statements:

- *"It is not funny, it is important".*
- *"I have never seen something similar".*
- *"It has advantages and disadvantages, is huge and it would be weird to have it at home if the person has cognitive impairment".*
- *"It seems to be useful and its working seems easy to learn".*
- *"It would be useful if it have times and places of working".*
- *"as a structure is not pretty".*
- *"One should learn how to use it"*
- *"It is very big for a house".*

Summarizing the opinions of the users regarding the system, it is useful if you are alone and cannot move, it is big for an apartment and they are worried about how to learn to use it, but they would be willing to learn to use it. Qualitative responses and comments of the participants seem to confirm that they would like to use the system *("I would use it if I would need it", "Everything new is nice. It will be helpful to bear stuff"*), although they would like to add more functionalities ("*Should bring the stuff nearer", "I would like it to feed me", "But will help me with personal hygene? No? Then I don't want it", "I would like it to help me to get up, my problem is that I cannot get up from the chair or the floor"*).

2.1.4 Discussion

The system is perceived as useful, as far as the person lives alone and is not fully autonomous. 69,23% of the participants reported positive or neutral responses to the question "Would you use the functionalities?", 83,33% reported positive or neutral responses to "The robotic arm can help me to achieve goals in my daily routine", 71,43% to "More control over my daily life", and 71,43 to "I would use this system when it becomes available".

The users were all relaxed in front of the robotic arm, and several users reacted to the arm showing curiosity for the system. Complementarily, the system is perceived as interesting, showing interest on technology. They express their concern about how to learn to use it, but they would be willing to use it.

Participants point to some difficulties regarding the size of the robotic arm and concerns about its cost.

2.2 VISUALIZATION TRIALS

In the visualization trials, we have collected data from professionals working in areas similar to those involved in SRS. In these trials, we have also combined ad-hoc questions with the AttrakDiff.

2.2.1 Participants

We recruited 13 professionals operators. 7 professionals operators worked for a tele-assistance company (SaludNova, http://www.saludnova.com/) and 6 for Matia - Ingema hospital. Employees from SaludNova were interviewed in the facilities of the company. The rest of the participants were assessed in the rehabilitation gymnasiums of the Hospital where we also conducted the Manipulation trials (see Figure M1).

7 woman and 6 men participated, all of them were medium-aged adults (mean = 31.08 years; st. dev. = 5.22). They did not present physical or sensorial problems hindering the performance of the planned tasks, apart from corrected myopia.

All of them have university graduates or equivalent (mean =22.09 years of education; st. dev. = 3.44) and have a job with medium to high responsibilities related to telecare and /or gerontology. All they reported to have jobs with high cognitive and social requirements, and low physical requirements.

Their experience with technology is high. They use all the devices included in the questionnaire (washing machine, dish washer, vacuum cleaner, TV, mobile phone / smart phone, PC / laptop), and they report to have been using PC / laptop a mean of 14.75 years (st. dev. = 4,495; min. 9 years, max. 21 years).

2.2.2 Usability, accessibility and learnability

Three tasks were established in order to let the participants use the system in a emulated natural scenarios of use (Figure 3). Usability of the system was evaluated by assessing the users' tasks performance regarding: task completion, errors, time, hints needed, comments, and usability difficulties observed.

Three specific scenarios were developed based on the general scenarios of the project, in order to allow the user interact with the system: grasping, leaving and moving scenario. Complementarily, ad-hoc quantitative questions were done about these issues after the three scenarios.

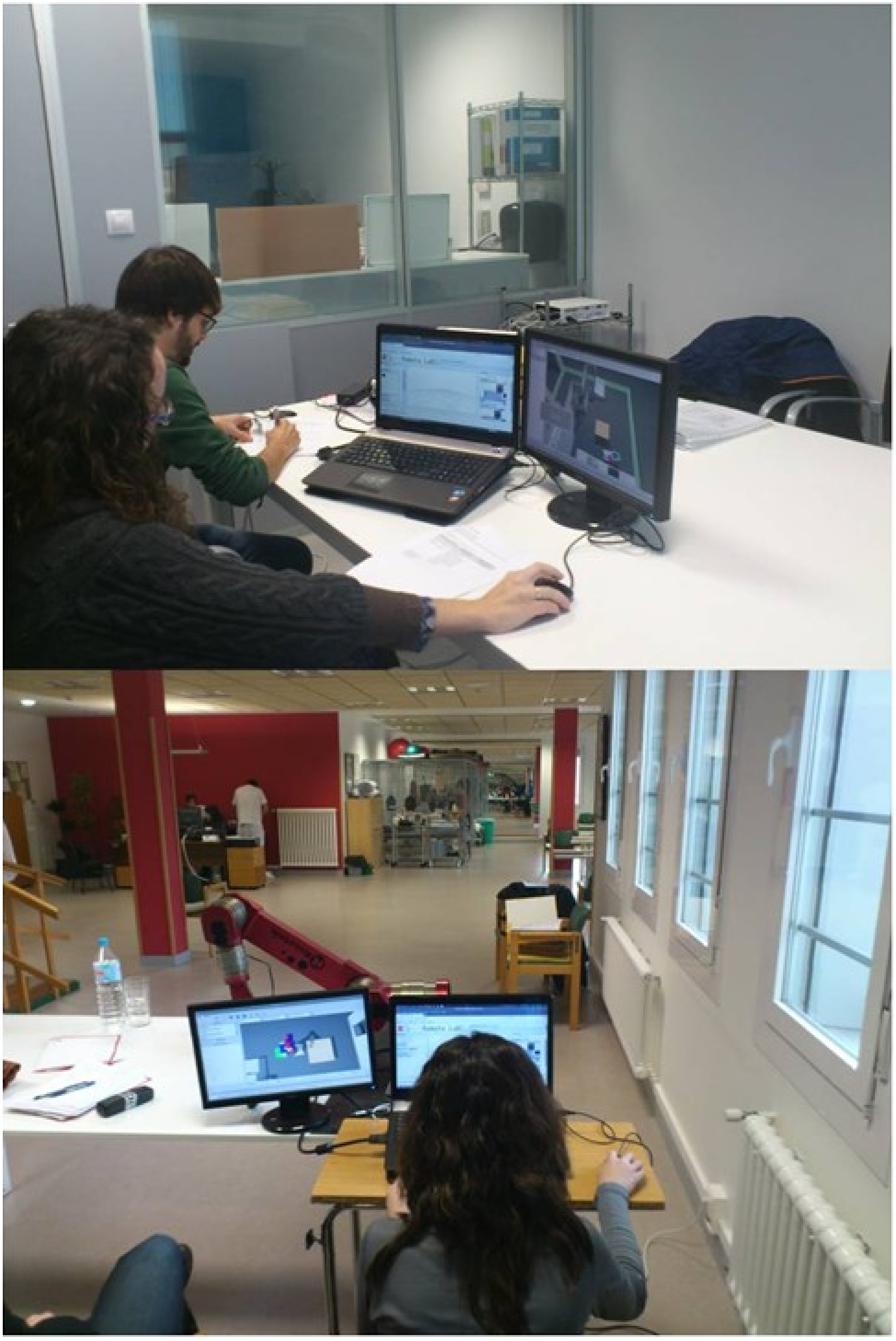


FIG. 3 VISUALIZATION TRIALS AT SALUDNOVA FACILITIES (ABOVE) AND MATIA HOSPITAL'S REHABILITATION GYMNASIUM (BELOW).

A) Grasping

Task Completion: 11 out of 13 completed the task.

Time: Mean= 92.16 seconds st. dev. = 33.9 seconds. The fastest of the users completed the task in 34 seconds meanwhile the slowest of the participants that completed the task needed 159 seconds to complete the task.

Errors: 0.53 errors on average were made by the users. Users press the grasping button previously to the robot achieve the correct grasping position.

Questions asked: 0.92 questions on average were made by the users. That means that nearly each user ask at least a question about the functionality or how to perform the task after having been explained.

Hints given: 1.38 hints were given on average. Nearly a third more hints that the users asked for.

Facial expression: The expression of the participants was neutral in most of the cases no one seemed to be frustrated, scarred or angry about using the system.

Comments (made by the user or the observer):

- The robot was out of place and as got near to the object it knocked out.
- *"How can I know if the robot is going to move or not? The software should notify of the movement status"*
- *"When should I press "grab"?"*
- The user changes the screen views, he takes more time but he customizes the view. States that he miss to know in which movement step is the robot.
- The user comments about the distance of the arm*. "The failure was that I pressed grab too early".*
- The user forgot to press the "show list" button. She pressed too early. She asked whether she has to tell the robot which object to grab. She suggests to place a marker that notifies the steps ending in order to be sure about the movement had ended.
- The hand was closed when the movement started so it was unable to grab the object.
- *"It should be more automatic. There is a lack of feedback, for example to know when grab".*
- The user presses grab before the arm is in position and asks if the arm is moving, if the system has received the command. The participant suggests to include feedback of the performance of the robot.

B) Leaving:

Task Completion: All the users completed the task.

Time: Mean= 125.69 seconds st. dev. = 44.57 seconds. The fastest of the users completed the task in 31 seconds meanwhile the slowest of the participants that completed the task needed 187 seconds to complete the task.

Errors: 0.38 errors on average.

Questions asked: 1.07 questions were asked on average, however st. dev. = 1.03 that means that there were several users that did not asked a question meanwhile other asked 2 or even 3 questions.

Hints given: 1 hint was given on average.

Facial expression: 11 out of 13 were neutral regarding the task, but 2 users expressed annoyance.

Comments (made by the user or the observer):

- The hand was closed at the start and did not open. The user asked about the system performance. After that it presses the grab button before the arm is in position.
- The hand was closed and knocked out the brick. After that the hand opened.
- The user asks again about the grabbing. He considers that the grabbing button should be in the same level to make the tasks sequential. He speaks also about the interface, he suggests to simplify and place the buttons in a more intuitive way. For example, "*the STOP buttons should not be there.*"
- The user was reminded of the need to move the arm.
- The user moves the arm too early. He suggests that the system should be more automatic, not to wait to the completion of the different steps.
- The user presses the grab button too early.
- The user takes long in move the hand and presses grab too early. He asks if he has skipped any step and gets surprised when the hand closes.
- It did not grab the object because the robot did not get near correctly. The user asked how to restart the system. He seems a little bit worried because he did not see the movement progress.
- The users suggest that when the hand grabs an object the content of the object should be taken into account to avoid spilling the content with the different movement performances options.

C) Moving scenario:

Task Completion: All of the participants completed the task.

Time: Mean= 218.15 seconds, st. dev. = 133.88 seconds. The fastest of the users completed the task in 89 seconds meanwhile the slowest of the participants that completed the task needed 567 seconds to complete the task.

Errors: 1.35 errors on average.

Questions asked: 1.38 on average

Hints given: 1.45 on average

Facial expression: The expression was neutral for most part of the users, just one user expressed frustration about controlling the robot.

Comments (made by the user or the observer):

- The user throws the milk brick before finishing the movement. "It's complicated to control the joystick looking at two different screens. It would be easier to control the robot using the cursors. The best would be to control the arm with a system that engages the robot arm with the arm of the remote controller. It should minimize the robot movements making it one single and integrated movement, and it would make easy not to wait for the different movement steps. It's also probable to find problems with the real map if a user changes the object position in the real life."

- "It's very difficult to control it. You should attend to two different screens. One to see the robot and the other to manage the joystick. It should have lateral movement in every axis in order to facilitate movements of 90º without having to make a turn.
- The user hits the table with the robot. States that he is not being able to move it where she wants. She asks where the robot is looking. She does not understand the joystick directions. *"It is not intuitive; it does not follows the robot space and orientation. The screen should turn when the robot turns".* She thinks that with cursors should be more difficult but better to make the more precise movements. She does not understand why there are 2 screens to control the robot. She asks how the acceleration works.
- *"The forward control is not intuitive. It should be some information about the robot orientation."* Due to the different screens, the attention is differed and makes complicated to control the robot. *"I don't know how far is the table from the robot"*
- The system fails and it is needed to restart the system. The user hit the oven and the table. The user states that it is easier to perform the movements if first the robot is oriented and afterwards starts the movement. "*The interface seems to be easy but I don't understand what the other buttons are."*
- The user releases the grabbing movement when she looks to the other screen. She has difficulties to differentiate between the robot directions when are not coinciding with the joystick ones. She states that it is very complicated to move the robot using the joystick. *"Probably with a robot image it would be more of guidance".*
- The brick falls to the ground; it was well placed but not assured with the robots hand. *"It's very messy to use the joystick, also the distance between the robots and the objects. It is not intuitive."* The participant asks how wide the robot's body is to avoid hit the furniture and how to control the speed. The participant suggest that the robot should move to the point that the proffesional points in the map or following a route that he draws in the map.
- *"It's very complicated!"* The participant fails moving backwards when should have moved forwards.
- *"It's easy to learn how to performance, despite sometimes you lose control"*
- The user asks why the robot is taking so long to react to the given order.

D) Quantitative questions

From the total sample of 13 participants, 12 participants answered the questions about the usability, accesibility and easy-to-learn of the system. About usability, most of the negative responses were focused in the feedback provided and the sense of control of the machine (Table 3), which must be improved in next versions of UI_PRO.

TAB. 3 REQUENCY OF PROFESSIONALS' RESPONSES TO USABILITY AND ACCESIBILITY ISSUES IN THE VISUALISATION TEST

	Definitely yes	Rather yes	Neutral	Rather not	Definitely not
Would you like to use the functionalities?	3	4	5		
I find the system easy to use.	1	8	2	1	

How well did you think the system show actual location of the robotic arm?			2		10
Is it easy to make the software do exactly what I want	1	4	4	1	2
I perceive perfect control over the activity of the system.	2	2	2	5	1

About learnability, most of the responses were positive (responses about easy-to-forget must be interpreted inversely) (Table 4), showing a good perception about how the system can be learned in spite of its usability problems.

TAB. 4 FREQUENCY OF PROFESSIONALS' RESPONSES TO LEARNABILITY ISSUES IN THE VISUALISATION TEST

	Definitely yes	Rather yes	Neutral	Rather not	Definitely not
To learn to use this system would be easy for me.	5	5	2		
It would be easy for me to learn skilful use of this system.	5	5	1	1	
It is easy to forget how to do things with this system.			2	8	2
Working with this system could be mentally stimulating.	4	1	6	1	

2.2.3 Acceptance

In this test, responses in the AttrakDiff are more widely distributed, especially in those adjectives more closely related to usability issues. Professionals agree that the technology is inventive and interesting, but also slow and not clearly meeting the expectations of the users (see Table 5).

TAB. 5 FREQUENCY OF PROFESSIONALS' RESPONSES IN THE ATTRAKDIFF FOR THE MANIPULATION TEST.

	1	2	3	4	5	6	7	
annoying	-	1	2	1	3	4	1	**enjoyable**
not understandable	1	4	-	3	4	-	-	**understandable**
creative	4	3	1	4	-	-	-	**dull**
easy to learn	2	4	2	2	2	-	-	**difficult to learn**
valuable	4	2	2	4	-	-	-	**inferior**

boring	-	-	2	4	-	5	1	**exiting**
not interesting	-	-	-	1	4	5	2	**interesting**
unpredictable	2	1	1	1	5	2	-	**predictable**
fast	1	-	-	3	3	1	3	**slow**
inventive	4	4	1	2	1	-	-	**conventional**
obstructive	1	1	-	3	3	1	3	**supportive**
good	3	3	2	4	-	-	-	**bad**
complicated	-	1	1	2	4	1	3	**easy**
unlikable	-	-	-	3	4	4	1	**pleasing**
usual	-	-	-	-	2	7	3	**leading edge**
pleasant	3	5	3	-	-	-	-	**unpleasant**
secure	1	4	2	2	1	1	1	**not secure**
motivating	3	4	3	1	-	1	-	**demotivating**
meets expectations	3	2	3	1	2	1	-	**does not meet expects.**
efficient	2	2	4	1	3	-	-	**inefficient**
clear	2	5	3	1	1	-	-	**confusing**
impractical	-	-	2	3	2	3	2	**practical**
organized	-	4	4	2	1	-	1	**cluttered**
attractive	2	4	2	2	2	-	-	**unattractive**
friendly	1	2	4	1	3	-	1	**unfriendly**
conservative	-	-	-	-	3	7	2	**innovative**

Regarding the acceptance ad-hoc questions (Table 6), we obtained:

- A higher percentage of participants with positive responses (Likely and Very likely) in items 1, 2, 4, 5, 8, showing that the professionals interviewed consider the robotic arm useful for the frail older adults and interesting for their work (Table N). 58% of the participants responded positively to the possibility of using the system when it becomes available, 42% gave neutral responses and no negative responses were collected.
- An equal percentage of participants with positive and negative responses to items 3, 6 and 7. As in the former section, results do not show relations between technology acceptance and system acceptance.

TAB. 6 FREQUENCY OF PROFESSIONALS' RESPONSES TO ACCEPTABILITY ISSUES IN THE MANIPULATION TEST

	Very unlikely	Unlikely	Neutral	Likely	Very likely
1. The robotic arm can help to help frail older adults with their daily routines	-	2	-	6	4
2. I perceive the use of the robotic arm can help me to control over the difficulties of the daily users.	3	2	1	4	2
3. This system would make my work easier.	3	1	4	2	2
4. This system would make my work more interesting.	-	1	4	4	3
5. I would use the system because I like to use such appliances.	3	-	3	4	2
6. I would use the system because these appliances are modern.	4	1	2	3	2
7. I would use the system to keep up with the newest technology.	5	-	2	2	3
8. In general, I would like to use this system when it becomes available	-	-	5	1	6

2.2.4 Differences between subgroups

Since the total sample in this study was composed by 7 persons from a tele-assistance company (SaludNova) and 6 persons from Matia – Ingema working with older adults but not directly involved in tele-assistance activities, statistical analysis were carried out in order to study differences between these subgroups.

T-Student comparison carried out with SPSS 18.0 showed similar results in both groups, with differences only showed in:

- Number of questions asked about the Grasping scenario (F=74,667; p<0,001), which mean where higher for SaludNova subgroup compared to Matia – Ingema subgroup.
- Perceptions about how organized – cluttered is the system (F=8,198; p<0,05), with SaludNova subgroup perceiving it as more organized.
- Perceptions about the easy-to-use of the system (F=6,664; p<0,05), about how well the system show actual location of the robotic arm (F=18,519; p<0,01), and about the intention to use the system because they like to use such applicances (F=6,542; p<0,05), which mean where higher Matia – Ingema subgroup compared to SaludNova subgroup.

Although these differences seems to show different level of knowledge about the interface tested, with SaludNova employees showing some higher degree of knowledge about the visualization system, we have decided to consider our results as a whole because the general agreement between these two subgroups in a wide range of usability, learnability and accessibility questions

2.2.5 Discussion

The UI-PRO should be improved in further versions in order to make the software usable and accessible. Responses about the user experience are more widely distributed compared to Manipulation test, especially in those adjectives more closely related to usability issues. Professionals agree that the technology is inventive and interesting, but also slow and not clearly meeting the expectations of the users. Negative responses about usability issues are focused in the feedback provided and the sense of control of the machine. Because of the difficulties found, we have constructed a table with design recommendations for UI Pro, including specific sections about grasping and leaving functionalities, movement of the robot and screen appearance (Table 7) and a suggestion of usable screen based on results collected (Figure 4).

TAB. 7 DESIGN SUGGESTIONS BASED ON VISUALIZATION TRIALS RESULTS

Usability and accesibility flaws	Design suggestions
Screen appearance	
1. Two different screens split attention while performing actions.	One screen to avoid divided attention / simplify multiple-screen solution.
2. The users focus their attention on environment-simulation screen and not on the movement pad screen, producing inaccurate movements.	
3. The robot-position map does not provide useful information to the user.	The environment simulation screen should give more detailed spatial information.
4. It is not possible to look at the joystick and the environment-simulation screen	The controls should not need to be visible in the screen, but a feedback on the movement.
5. There is no screen-feedback about the actions performed in the robot-movement plan	Feedback is needed across the different steps
6. There are several buttons and options that are not needed or functional in the screen.	There is no need of additional stimulation if not functional: i. E. “The label Care’o’bot Remote Lab” label should be removed.
7. The interface is not intuitive.	The buttons should be placed in relevant places in the screen. All the actions of the same level at the same place, icons instead of buttons, text boxes or windows with explanations, etc.
8. The Stop button should be bigger in case of emergency.	
Movement of the robot	
9. Is difficult to understand where the robot is located when starting a movement.	The robot direction should be shown in the screen whenever the user is going to start a movement sequence.
10. It is difficult to know the distance between the robot and the objects.	Apart from collision-avoidance systems, the distances should be constantly notified to the user to prevent collisions.
11. It is more difficult to move the robot compared	The system could be moved just by pointing the end

to pointing the direction the robot move.	direction in the map.
12. The environment simulation upper point of view seems to be difficult to understand than the subjective point of view	The system should provide subjective vision
13. The users prefer to minimize the action options	It would be preferable if the system detects the object automatically than generate a showlist.
	The system could allow to point and click one object in the environment simulation screen.
14. Users have problems to move the robot in "L" movements	Lateral movements could be promoted, instead of turning.
15. The movement stops when the user grabs the pointer out of the square movement pad	The robot movement should not stop if the user got off the square movement pad
16. It's difficult to know how to speed up or slow down the movement robot.	There should be an speedometer or acoustic signals (i.e. high frequency sound for higher speeds) to inform about the speed of the robot
17. The joystick directions not always coincide with the robot orientation in the environment-simulation screen.	The environment-simulation screen should always move with the robot orientation
	Moving the robot could be done by cursors instead of joystick simulation interface
Grasping and leaving functionality	
18. The user does not know when to press "grab".	There should be feedback about the different grabbing steps. The user should be aware of the precise moment when pressing the grab option; the system should not allow the user to start a grabbing action which outcome is clearly a fail.
19. It is difficult to know in which arm movement step is the robot.	There should be some visual feedback about the movements.
20. The users press the grab button too early knocking off the object.	When the hand is closed in advance by error the robot movement should be stopped automatically.
	The hand should be able to be opened again.
21. The robot takes too much time to start performing the action.	The system should give feedback to the user about the robot starts to perform actions. "Please wait" message.
22. The movement performance of the hand does not take into account the object (the opened tins content could be spilled)	The movement should take into account the object automatically, or ask the user for movement safety degrees.
23. Feedback is needed in the arm movement sequence.	
24. There are unnecessary options in the menu.	The actions menu is not intuitive but learnable. Reduce the options, automatize up to the functionality. The users should skip the steps as much as possible and only choose on functionalities.

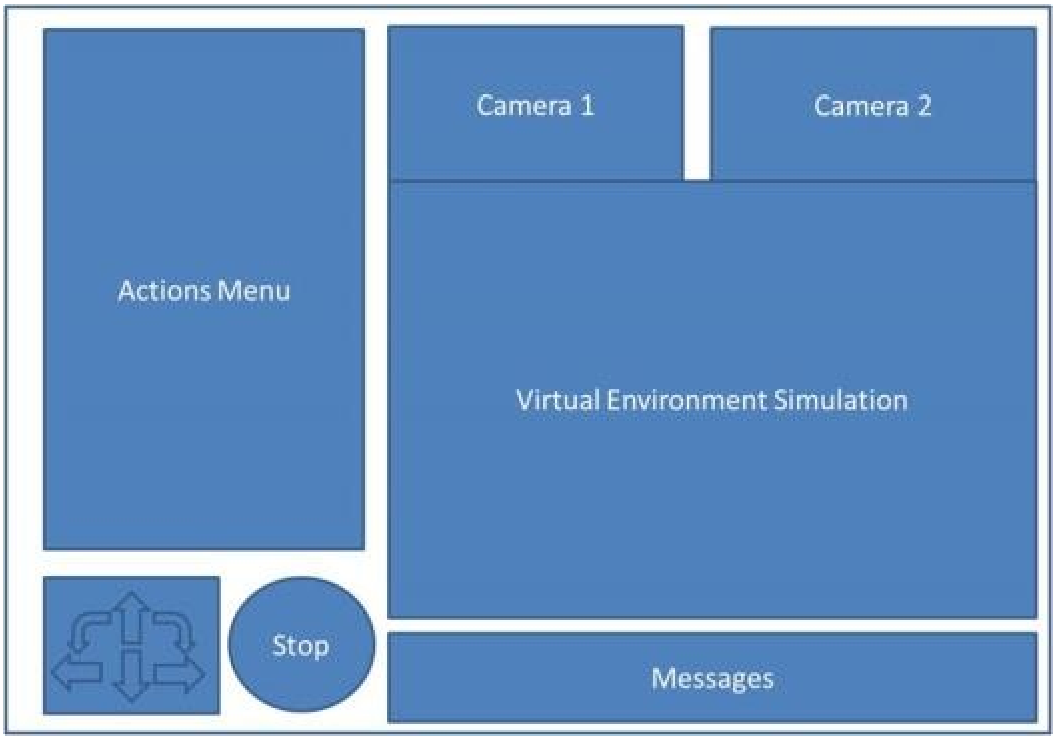


FIG. 4 SUGGESTION OF USABLE SCREEN

Apart from usability and accessibility results of the UI-PRO, these trials have showed that the user acceptance of the interaction with the robotic arm through an interface like the one presented is medium to high. In spite of this difficulties, all the participants provided positive or neutral responses to *"I would like to use this system when it becomes available"*, 83,33% provided positive or neutral responses to *"The robotic arm can help to help frail older adults with their daily routines",* 91,67% to *"This system would make my work more interesting"* and 75% to "*I would use the system because I like to use such appliances".* About learnability, most of the responses were also positive showing a good perception about how the system can be learned.

3 WHOLE-SYSTEM PRE -TEST IN REAL HOME

3.1 TEST OBJECTIVES

The main objectives of this test were to identify technical problems when operating the robot in a real home environment and to get a first feedback from elderly users on their perception of the robot in their home.

3.2 TEST ENVIRONMENT AND PROCEDURE

The test was carried out in two apartments of the same house located near Stuttgart, Germany. Two elderly people (1 female, age 80; 1 male, age 81) live on the ground floor where the robot was deployed. For UI_PRI operation, a remote operator (grandchild, female, age 30) controlled the robot from another apartment upstairs in the same house.

The test duration was 1 ½ days including setup and removal of robot and equipment. There were two test trials with elderly people, one involving a remote operator. Two interviewers were attending to the participants, corresponding to two usage scenarios:

- Scenario 1: An elderly person was sitting on the couch and used a handheld interaction device to make the robot fetch a book from a locker in the dining room. However, the robot failed at executing the task (this failure was planned / simulated) because a stool hindered appropriate path planning for delivering the object to the user. Therefore, a remote operator (located in another apartment) was called and remotely navigated the robot to deliver the book.
- Scenario 2: An elderly person was in bed and the robot fetched a medicine box from the window sill in the kitchen.

The following evaluation methods were used:

- Evaluation list for technical performance of system components
- Interviews using prepared questions on robot perception

In cases where system components failed or were unavailable, a wizard-of-oz approach was adopted and the robot was remotely controlled with a joystick in order to still be able to obtain user opinions on the fulfilled scenarios. Autonomous and semi-autonomous operation was resumed after resolving the error state.

The sessions were recorded on video with four cameras installed in the apartment and photos were taken.

In the following, pictures of the test environment are shown with the robot in action (figures 5 to 12).

FIG. 5 EXTERIOR OF HOUSE WITH ROBOT ENTERING (LEFT); CORRIDOR WITH ROBOT IN OPERATION DURING TEST (RIGHT)

FIG. 6 MOST NARROW PASSAGE IN THE APARTMENT (BETWEEN DINING ROOM AND KITCHEN-LEFT) WITHOUT ROBOT; WITH ROBOT PASSING (COULD NOT BE ACCOMPLISHED IN AUTONOMOUS PATH PLANNING MODE-CENTRE); LIVING ROOM-RIGHT

FIG. 7 BEDROOM; PASSAGE BETWEEN LIVING ROOM AND DINING ROOM

FIG. 8 BOOK TO BE GRASPED ON LOCKER

FIG. 9 MEDICINE BOX TO BE GRASPED ON KITCHEN WINDOW SILL

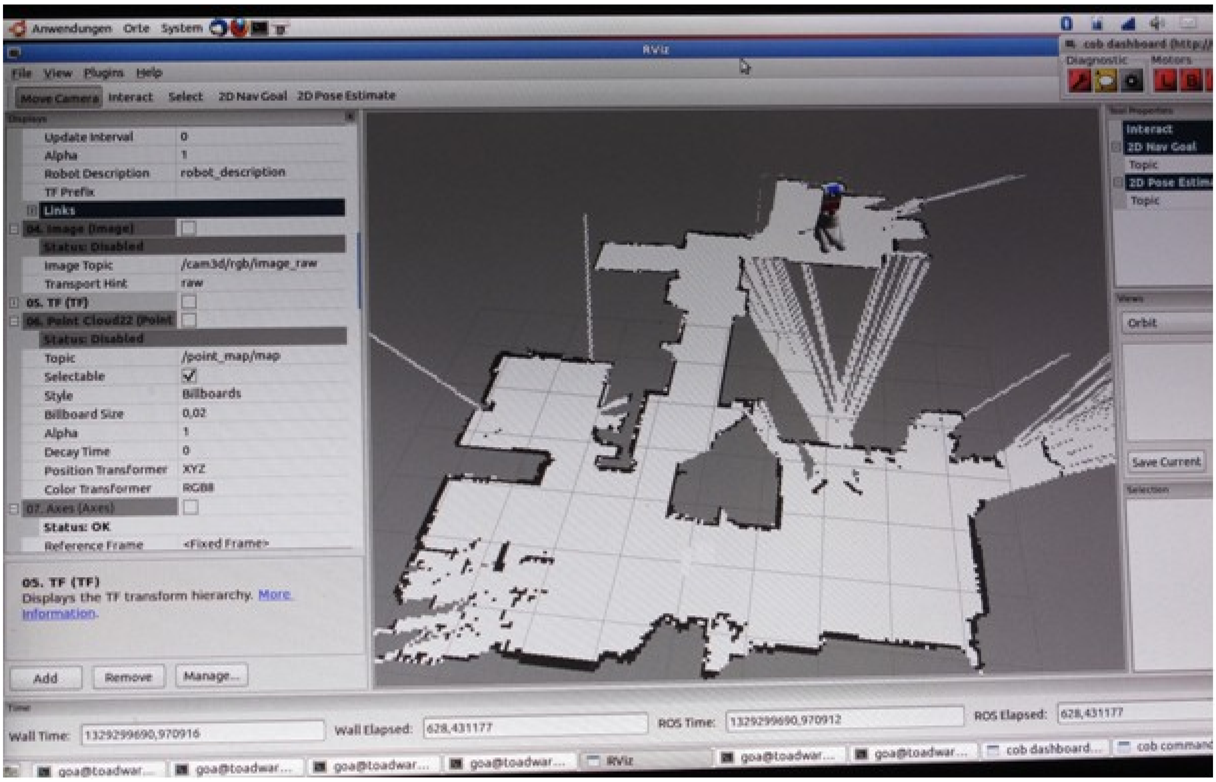


FIG. 10 FLOOR PLAN GENERATED FROM SENSOR DATA BY THE ROBOT MOVING AROUND IN APARTMENT (FRONT LEFT: DINING ROOM; FRONT RIGHT: LIVING ROOM; CENTER LEFT: KITCHEN; BACK RIGHT, WHERE ROBOT IS LOCATED: BEDROOM)

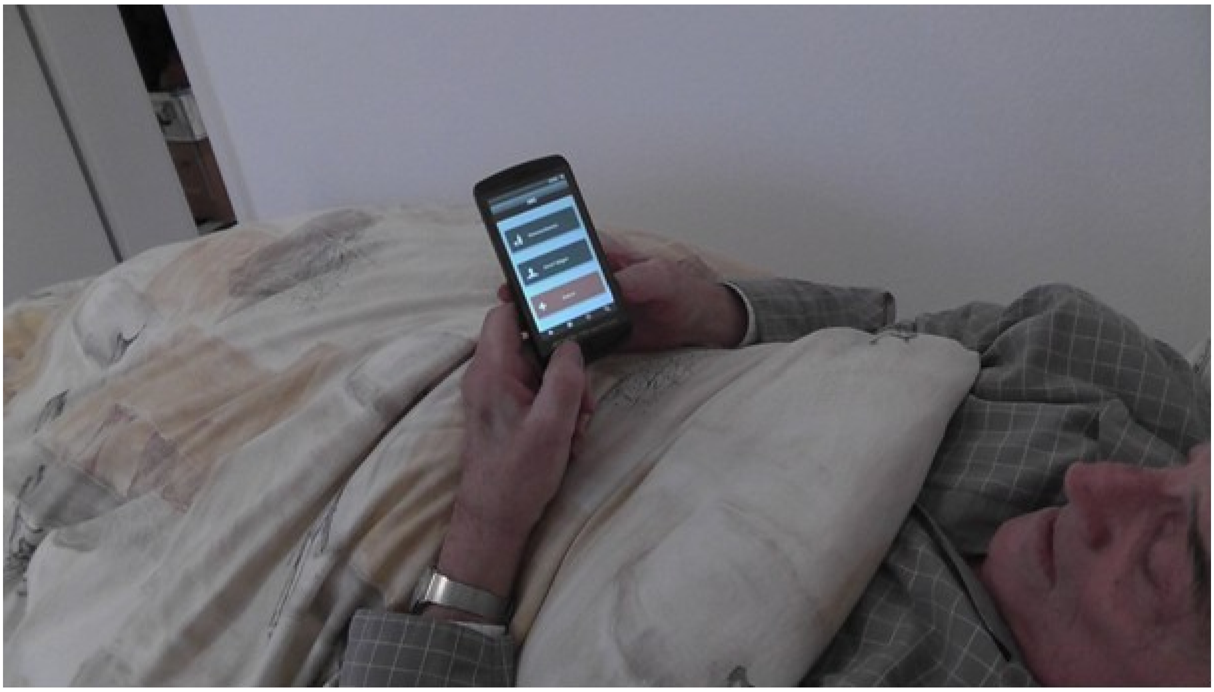


FIG. 11 ELDERLY PERSON OPERATING UI_LOC (HANDHELD ANDROID DEVICE) IN BED DURING MEDICINE FETCHING SCENARIO

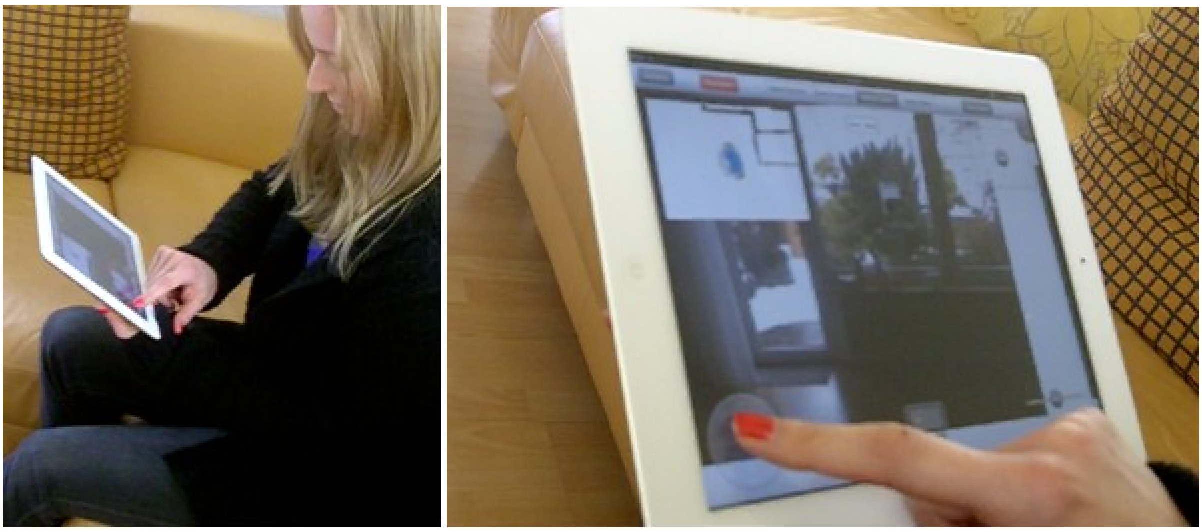


FIG. 12 GRANDDAUGHTER CONTROLLING THE ROBOT FROM A REMOTE SITE USING UI_PRI (IPAD)

3.3 RESULTS

3.3.1 Technical problems

1. The most significant problem was that the current state of the SRS components did not allow for an operation close to the anticipated operation. For a large part, it was necessary to improvise by using wizard-of-oz (tele-operated mode).
2. UI_LOC was not usable because it is not connected to the SRS system yet.
3. Generation of map for UI_PRI usage did not work; therefore no navigation by map possible
4. Problem operating the "turn" control in UI_PRI's manual mode: It was often overlaid with the view button (see figure 13). The only way to recover was to kill the app.
5. Remote user did not succeed in navigating the robot to the desired position using manual control mode in UI_PRI (likely because of field of view issues and problems estimating position and distances)
6. Narrow passages like doors were a significant problem. Door sizes in the apartment were between 73 cm and 80 cm. The robot was unable to pass any of them in autonomous navigation mode. In fact, passages did not even have to be very narrow. The robot also initially could not find a path through a wider passage between the living room and dining room (see figure 14). Only after we removed the plant on the left and used another mode of path planning was the robot able to navigate through the passage.
7. Quite a lot of space is required around each side of the robot for successful navigation and manipulation activities. For example, 1 meter on the right side for the arm to move and 0.8m for the robot to turn around its own axis.
8. It was not feasible to teach the two test objects (book and medicine box); therefore an already-known object (milk box) had to be used for object detection and grasp planning which was then quickly exchanged manually before grasping execution (see figure 15).
9. During grasping, the medicine box was slightly crushed and the book cover became damaged because the book bumped against the locker (these issues might have occurred because a substitute object, the milk box, was used for calculating the grasping configuration, rather than the real objects to be grasped)

10. Lighting: The robot was unable to recognize objects on a window sill due to backlight (light entering through the window from outside). It was necessary to close the shutter and use artificial indoor light to achieve successful object recognition (see figure 15).
11. Grasping objects in corners of the apartment or in cramped places was impossible
12. The robot was unable to pick up a book lying flat on a locker. The book had to be standing in upright position (see figure 16, left).
13. The robot was unable to pick up a medicine box lying on a window sill if the box was too close to the window (gripper fingers have to grasp behind object). It had to be positioned on the front side of the sill facing the robot (see figure 16, right).
14. While the robot was moving , the book fell down from the tray along the way because the only way it could be placed on the tray was standing vertically
15. Robot hardware: When passing door sills, the outer shell of the robot often touched the sill. Crossing sills in autonomous mode was not possible. It could be beneficial to leave more space below the lowest point of the robot shell in a future revision of the robot (see figure 17).
16. The battery had to be recharged after around 4 hours of operation (unlikely to be a problem under normal usage conditions)

3.3.2 Necessary changes to the apartment

1. To create the map of the apartment, some furniture and curtains had to be removed temporarily only.
2. All carpets had to be removed permanently so the robot could drive properly.
3. To grasp a book from a cupboard, an adjacent table had to be moved significantly so the robot had enough space for operation (see figure 18).

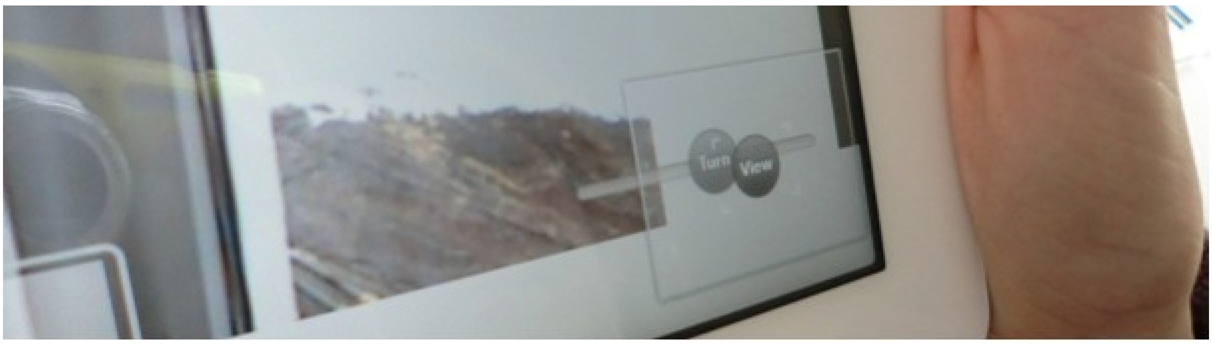


FIG. 13 PROBLEM WITH “TURN” CONTROL BECOMING UNUSABLE

FIG. 14 PASSAGE CAUSING PROBLEMS FOR PATH PLANNING

FIG. 15 MILK BOX (ON SILL IN THE LEFT PICTURE) AS A SUBSTITUTE OBJECT FOR MEDICINE BOX (ON SILL IN THE RIGHT PICTURE) TO CIRCUMVENT TEACHING OF MEDICINE BOX; USAGE OF ARTIFICIAL LIGHT TO ENABLE OBJECT DETECTION BECAUSE OF BACKLIGHT ENTERING THROUGH WINDOW

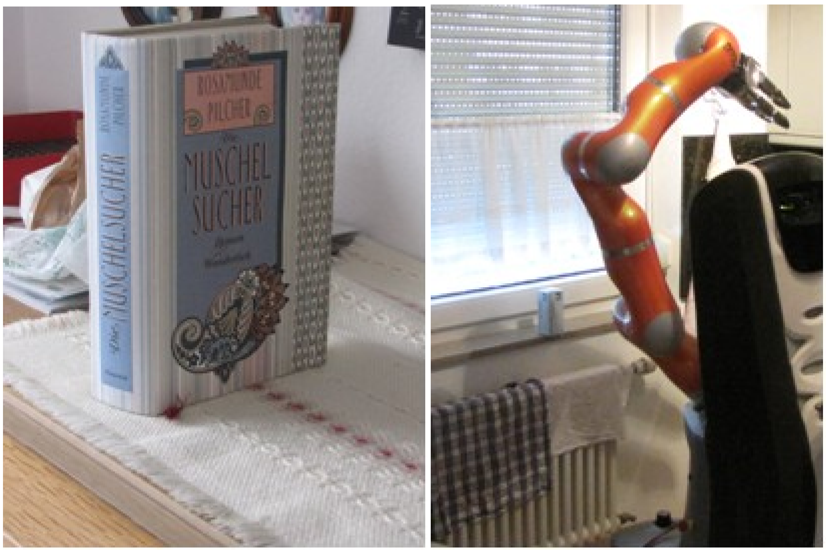


FIG. 16 MILK BOX (ON SILL IN THE LEFT PICTURE) AS A SUBSTITUTE OBJECT FOR MEDICINE BOX (ON SILL IN THE RIGHT PICTURE) TO CIRCUMVENT TEACHING OF MEDICINE BOX; USAGE OF ARTIFICIAL LIGHT TO ENABLE OBJECT DETECTION BECAUSE OF BACKLIGHT ENTERING THROUGH WINDOW

FIG. 17 PROBLEMATIC DOOR SILL (LEFT); LOWER BOUND OF ROBOT SHELL (RIGHT)

FIG. 18 TABLE HAD TO BE MOVED TO THE LEFT FOR ROBOT TO GRASP THE BOOK ON THE LOCKER (LOTS OF SPACE REQUIRED AROUND ROBOT)

3.3.3 User perception of the robot

In general the two elderly test users felt safe with the robot and had trust in it. The following statements were made:

Robot appearance: It was stated that the robot looks very *"friendly"* and *"sympathetic".* They had imagined it to be much larger (which would have been negative) and much more *"industrial".* The design of the robot was described as *"very good"*. A stated reason was that the technology is hidden and the robot looks like a home product. It was stated that it "*fits into an apartment like an armchair".* The round contours and soft outer shell were mentioned positively. The robot's size was perceived as *"good".* It should be noted however, that the test took place in a rather large apartment. Robot size would likely be more critical in smaller apartments.

Robot movement and sound: Operation was found to be very quiet, which was perceived positively (the fan of the built-in computer was the loudest sound emitted). Robot movement was perceived as slow and smooth. For this reason, participants did not perceive the robot does as dangerous. Before the robot came to the apartment, the users had the idea of an industrial robot working fast, hectic, rough, and loud. Therefore, they were surprised in a very positive way. Wheeled operation was preferred over the possibility of a biped robot: *"It's nice that it's on wheels and not walking unnaturally."*

Robot safety: When asked about their impression of robot safety, the participants both agreed to "*feel safe*". When asked why, it was mentioned that *"I saw that it stops in front of an obstacle because of its sensors."* Also, the wheeled drive of the robot made an impression of being robust and therefore safe.

Robot scenarios: Once again, this evaluation showed that usage scenarios are very important for user acceptance. Users often reflected upon possible application scenarios. For example, it was mentioned that it would be helpful if the robot could bring some food when lying in bed because of illness.

4 ADVANCED SRS PROTOTYPE TESTS

The first comprehensive set of tests has been performed in the Milan apartment, focusing on the first evaluation of the whole prototype in a real home environment and with potential users of the system. The evaluation focused both on technical effectiveness and on SRS usefulness with respect to user peculiar requiremets, usability, learnability, and acceptance expressed by potential users (see table 8).

TAB. 8 DETAILS OF MAIN REASEARCH GOAL ASSESSED

Main research goals	Complete list of research questions
Technical Effectiveness	*evaluating success of each single task execution*
	Evaluating time needed to complete the tasks
	describing eventual problems occurred in tasks completions
SRS peculiar requirements	*Efficiency in ADL task completion*
	Secure in task completion
	Improving the autonomy
	Improving of communication and interaction modality
	Improving safety
	Easy Integration in the private home
Usability/ learnability	*Easy to Learn*
	Comfort perception
Acceptability/ intention to adopt	*advantages/disadvantages perceived*
	Acceptable from Psychological/emotional point of view
	Safeguarding of Privacy and Ethics
	Usefulness
	Attractiveness
	Eligibility (Intention to adopt)

4.1 METHODS AND OBJECTIVES

Detailed Methods of investigation were already reported in D6.1-b; however, it was necessary to make some changes to the designed plan, so table 9 reports here in brief the methodology actually adopted.

In particular, due to many technical problems, the scenarios designed in the course of the project and illustrated in D6.1-b were no possible to be tested.

As result of these limits, only one very simple scenario was designed after the three days of integration meeting in Milan (2-4/05/2012). The scenario can be expressed as "*autonomous fetch and bring of the medicine from the shelf to the sofa*" (figure 19). Videos showing tests executions are uploaded into srs website www.srs-project.eu, at the section gallery.

TAB. 9 SUMMARY OF THE METHODOLOGY ADOPTED TO TEST THE SRS PROTOTYPE

People assessed	• 16 elderly, 1 young disable man • 12 private caregivers , • 5 professionals operators
Test sessions	• **Before robot coming to Milan (last week of April)** -One our explanation of srs concept, showing videos and interfaces -Informed consent reading and signing • **4 test sessions each day with the robot in the test site (second week of May**) -Short introduction to participants -Up to 3 SRS usage experiences (depending on technical robot performance) -Questionnaires administration -Technical effectiveness recording
Indicators:	• **Technical indicators** -Evaluation check list for technical performance • **Users indicators** -Interactive think-aloud with moderators next to participants -Ad hoc developed quantitative and qualitative questionnaires -Standard Attrackdiff questionnaire -Safety-ethical-privacy issues focus group with moderators at the end of the test sessions
Users' validation if:	• Attrakdiff results→ results belonging to area called "desired" • Ad hoc questionnaires results→ indicators threshold ">3,5 (1 to 5 scale) • Qualitative consideration emerging, behavior observations and focus groups confirming quantitative answers

By the way, to test all the user devices (UI-LOC, UI-PRI and UI-PRO) and to assess the major number of user requirements, the same scenario was tested in three different situations showing to participants different objectives. They were the following:

Objective 1: Elderly improving their Autonomy at home

Elderly people alone at home, sitting on the sofa using UI-LOC autonomously send the robot to bring a medicine box located on a shelf in the corridor.

Objective 2: Private caregivers monitoring situation and remotely assisting

Private caregiver calls the assisted elderly people to check the situation at home. He/she realizes that the assisted forgot to take the medicine. The elderly accepts the help of the caregiver. The caregiver using UI_PRI sends the robot to bring a medicine box located on a shelf in the corridor, in the meantime the elderly and the caregiver keep talking, and the caregiver got also visual feedback of what the robot is doing and were it is going.

Objective 3: Professional operators managing emergency

Elderly feeling bed presses the emergency button. The professional operator is immediately contacted. The elderly states that immediately needs it's medicine. The professional operator using UI_PRO sends the robot to bring a medicine box located on a shelf in the corridor, in the meantime the elderly and the professional remote operator keep talking, and the professional operator has also feedback through the robot simulator of the robot movements and robotic arm.

FIG. 19 THE SCENARIO "AUTONOMOUS FETCH AND BRING OF THE MEDICINE FROM THE SHELF TO THE SOFA".

4.2 USERS RESULTS

4.2.1 ELDERLY PEOPLE RESULTS

16 elderly people were recruited. They were mostly people already needing some help at home at the moment, but still with good cognitive level (it was tested before the robot arrival their ability to understand SRS concept, to perform tasks requested and to answer to the questionnaires). So they presented mainly impairments related to mobility, low vision and or hearing problems, or simply handling difficulties due for example to Parkinson, or just slowness and stiffness in movements. Difficult in activities of daily living was assessed through Barthel index. Some of these people were recruited selecting people attending Don Gnocchi daily center (Palazzolo Institute, Milan). A center for those elderly who still live at their home but having some difficulties at home, fearing to stay all day long at home or feeling lonely all day long at home; spend some days of the week at the center were they receive lunch, and play some activities together with other elderly people from 10 a.m. to 4 pm. Other elderly people where recruited between people coming at the moment to the Don Gnocchi hospital to receive some rehabilitation treatments.

However in the sample recruited there were also some “young and quite fit” elderly people, so people may be needing help in the future.

We thought it was important to make trying the system both:

- to “*frail elderly*”, so people that already need for such kind of assistance, but in the most of cases very old so very little used to technology.
- to *“young elderly”,* because they are more used to technology, and because they will be more probably the future user of the system.

Table 10 and figure 20 report and show main features of the sample recruited.

TAB. 10 MAIN FEATURES OF ELDERLY PEOPLE RECRUITED

Main features of elderly people recruited	
Age	Mean 79,6 (st. dev. = 7,05) 71 min , 96 max
Gender	11 female and 5 male
Education (0 null, 1 first, 2 low secondary, 3 high secondary, 4 degree)	Mean 2,3(st. dev. = 1,34) 0 min , 4 max
Activity of daily living Frailty index (Barthel index: 100=completely independent in ADL)	Mean: 90,8 (st. dev. = 13,09) 55 min , 100 max

FIG. 20 SOME OF THE PEOPLE RECRUITED USING THE INTERFACE UI-LOC DURING SCENARIOS EXECUTIONS.

One young disable man (age 23, Barthel index: 33/100, education: 3) has also been recruited, to investigate other exploitation possibilities of the SRS system (figure 21). Results related to him will be reported separately.

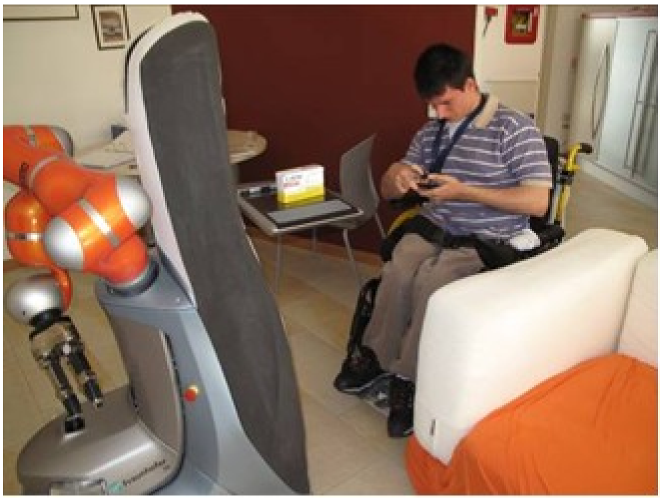


FIG. 21 A YOUNG DISABLE MAN RECRUITED TO TEST SRS PROTOTYPE

A) Ad HoC Questionnaire results - elderly people

The ad hoc questions were both quantitative (1 to 5 Likert type scale, ranging from "completely disagree" to "completely agree", success is considered reach for each question if mean elderly value is > 3,5); and qualitative; so asking "why" after the administration of each quantitative question. Each question was related to a particular research question (see table 8 or for more details D6.1-b) .

Qualitative results from why questions partially confirm quantitative results, however in some cases they are in conflict with quantitative results.

Figure 22 shows quantitative results (mean values and error bars) related to the questions addressed by the 16 participants.

Moreover, despite the hypothesis, no significant differences were found (Wilcoxon-Mann-Whitney Test [Wessa, P.; 2012], p values> 0,05 for every indicator), comparing the group of elderly people characterized by Barthel index =100, the so called "*young elderly*"; with the group of people characterized by Barthel index <100, the so called "*frail elderly*". Only differences related to usability with respect to the user's peculiar disability were observed; e.g. difficulties in using the UI-LOC for a man with hearing difficulties were different from difficulties of another affected by low vision. In the following part of the text these results are better detailed.

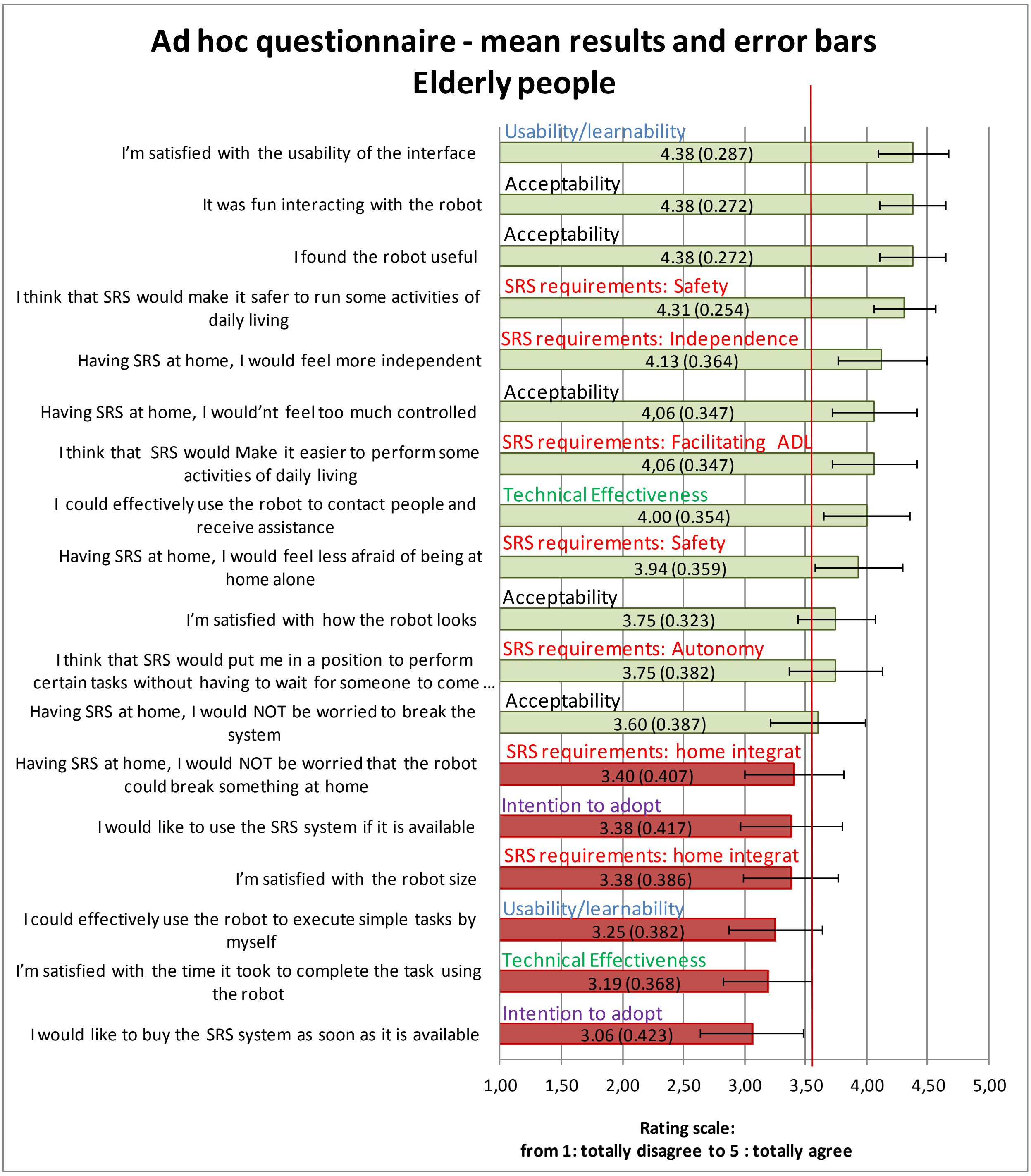


FIG. 22 QUANTITATIVE RESULTS OF AD HOC QUESTIONNAIRE (MEAN VALUES AND ERROR BARS) RELATED TO THE 16 PARTICIPANTS. SINGLE INDICATORS: GREEN IF SUCCESSFUL RESULT, RED IF NOT.

The majority of the elderly interviewed liked very much the idea of a robot helping them at home, with the possibility to be remotely monitored and supported by family or professionals operators. In general they are not scared (even if few of them were, mostly because of the arm), are curious and amused by the novelty and they found the usage experience very funny and interesting.

However they were well aware that the presented robot was only a prototype, not still completed and ready for market. Maybe for this reason, a lot of differences between quantitative and qualitative answers were found.

For example, asking if "is it the robot useful?"; the quantitative answers were often very good, scoring between 4 and 5 while qualitative answers were often sounding like "*well, of course it will be useful when it will be ready, because it will become able to bring a lot of objects and to perform a lot of tasks, isn't it?*" or, reflecting a common need between elderly people: "*It just take that medicine but I have to take many different medicines at different times, the robot would be able to recognize all of them? And also bringing me a glass of water to bring the medicines?"* or *"If it is just able to bring simple objects, well I would do it by myself, if it will be able to make more complex tasks like cleaning the house, handling and managing utensils...so it would be very useful*." The function for which they already really appreciated the usefulness was the possibility to communicate with their relatives (or with a professional operator), knowing that relatives were also seeing the situation at their home, with the possibility to give help. This function was very reassuring for them, nobody was worried about privacy, this option was just making them feeling reassured to be monitored in case of emergency. However only voice as feedback was considered not sufficient by someone (moreover people with hearing problems). Some of them stated that they also would have liked to see their relatives, for example directly on the robot coming to them in case they didn't hear the device ringing.

Another important point to be stressed is the usability of the user interface; the concept was considered simple and well understandable, consisting only of three big buttons, and few menus; so, at the question "are you satisfied with usability of the interface?", the quantitative answers scored 4,4, but at the question" I could effectively use the robot to execute simple tasks by myself" the quantitative question scored only 3.25. The qualitative answers did confirm this last result. Most of people said *"It's easy to understand but I would not be able to use it by myself".*

This aspect is primarily confirmed objective observations and by videos showing elderly people in real difficulties using a touch screen device; not only for problems related to the SRS app comprehension but for problems related to the complexity of the selected hardware (touch screen Samsung Galaxy). The test moderators had always to support them to finalize the order, or to make the call, or give them advice that the phone was ringing and they had to answer (because some of them hev hering difficulties).

Finally, three participants were illiterate, so they were not able to control the robot by themselves (reading the menus related to tasks), but they were able to learn how to contact remote operators (second button, then contacts represented by photographs) or to call for an emergency (red button, then green button to confirm).

Table11 reports the complete list of Usability problems emerged and some ideas to solve them.

TAB. 11 USABILITY PROBLEMS EMERGED WITH UI-LOC

Usability problems of UI_LOC
Insufficient, prolonged or inaccurate pressure on the touch keys. • Traditional button at least for emergency? • Voice control addition?
Too low call volume and speaker volume for many elderly having hearing problems • Adding vibration option? • Voice of remote operator coming from the robot?
Font size not big enough for some of the elderly people with vision problems • Vocal menus and "screen reader"?
Illiterate people can't read text • Adding icons wherever it's possible?
On the Galaxy device, when you receive a call, you have to activate the "speaker" (a small icon on the phone) an extra and complex operation.
"Go back" button is developed inside the UI_LOC app, but on the Galaxy device (an invisible icon on the phone, that appears only when you passes the finger on) • Back button integration on the UI_LOC software?
Skype not well integrated into the UI_LOC app. When you make a call, the calling (including emergency call) doesn't start autonomously, but it opens the Skype contacts, and you have to make a long sequence of actions to reselect the right contact and then finally make the call

Other qualitative results are instead coherent with quantitative results; they are related to the following issues:

- the robot is too big for their apartment.
- the robot is too slow, it takes too much time to complete a task, and many of them said *"imagine if I really need for that medicine"*
- the arm makes strange movements to bring objects, it rotates the object in the space, so many of them said the same thing at this regard: "*imagine if it was a glass full of water!*"
- Robot aspect as in the previous study, has split people in two parts, people desiring a machine like aspect, so happy about the current aspect, and people instead desiring a human aspect and a more human behavior.
- People fear is not related to the robot itself but about its possible malfunctions, the possibility that the robot break something inside the house, themself breaking the interface device or simply not being able to deal with the interface.

B) Attrakdiff results - elderly people

As better specified in D6.1, AttrakDiff™ is an instrument for measuring the attractiveness of interactive products. With the help of pairs of opposite adjectives, users (or potential users) can indicate their perception of the product. The following figures (23, 24, 25) report ATTRAKTIFF results obtained from elderly users interviewed in relation to user interface UI-LOC.

Figure 24 in particular show average values of the 16 elderly people, in the field of:

- **Pragmatic Quality (PQ):** usability of a product and indicates how successfully users are in achieving their goals using the product.
- **Hedonic Quality - Identity (HQ-I):** to what extent the product allows the user to identify with it.
- **Hedonic quality - Stimulation (HQ-S**): what extent the product can support needs in terms of novel, interesting, and stimulating functions, contents, and interaction- and presentation-styles.
- **Attractiveness (ATT):** global value of the product based on the quality perception.

Hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

The last figure (25) shows that the user interface was rated as "rather desired". The classification here is not clearly "pragmatic" because the confidence interval overlaps into the neighboring character zone. The user is assisted by the product, however the value of pragmatic quality only reaches the average values. Consequently there is room for improvement in terms of usability. In terms of hedonic quality the character classification does clearly not apply because the confidence interval spills out over the character zone. The user is stimulated by this product, however the hedonic value is only average. Room for improvement exists also in terms of hedonic quality.

The confidence rectangle is small. The users are at one in their evaluation of the product.

The users are at one in their ratings of both dimensions.

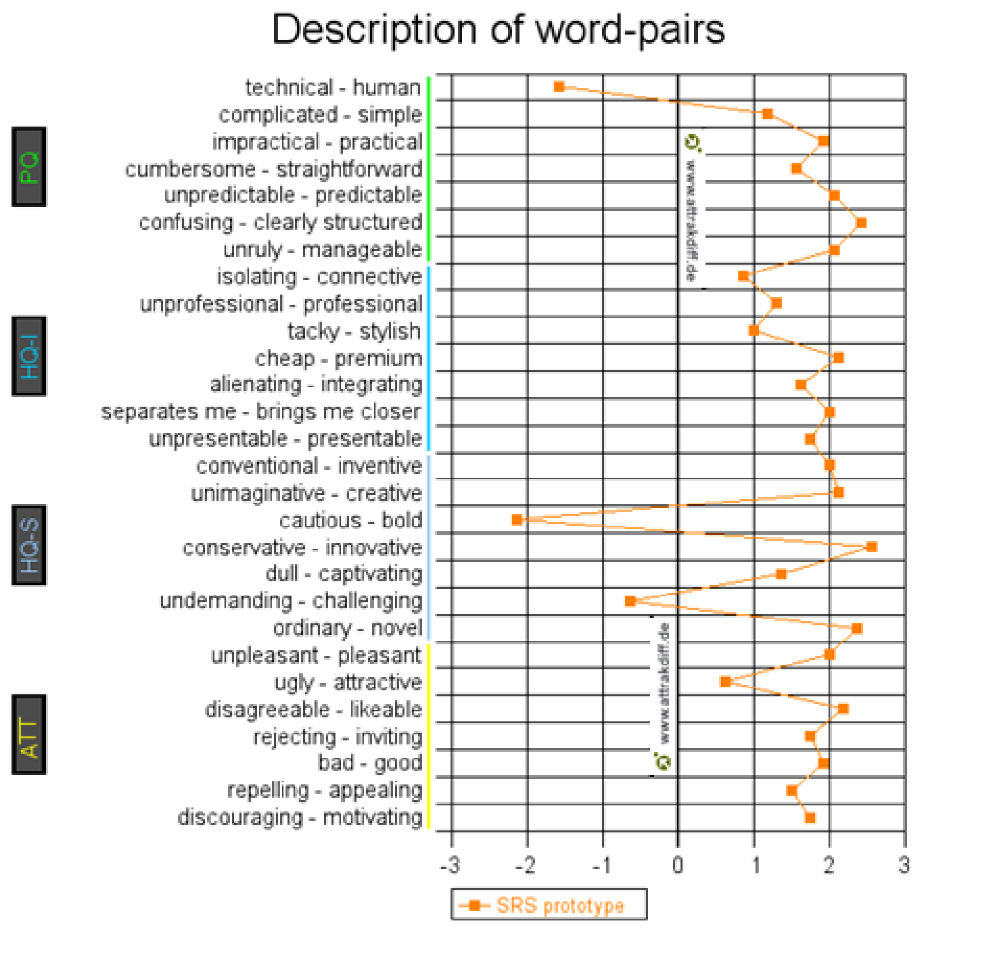


FIG. 23 ATTRAKDIFF RESULTS - DESCRIPRION OF WORD- PAIRS; MEAN VALUES - 16 ELDERLY PEOPLE

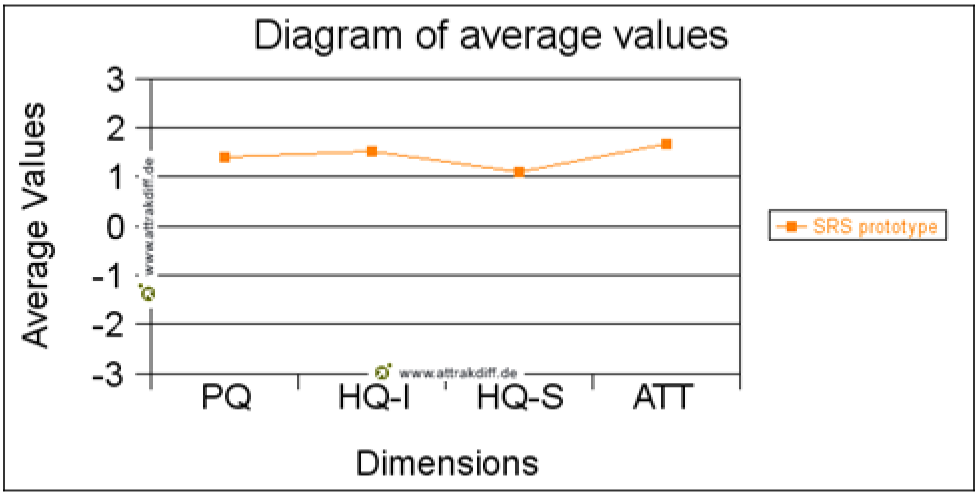


FIG. 24 DIAGRAM OF AVERAGE VALUES-ELDERLY PEOPLE

FIG. 25 PORTFOLIO WITH AVERAGE VALUES (ORANGE SQUARE REPORTING LETTER P) OF THE DIMENSIONS PQ AND HQ AND THE CONFIDENCE RECTANGLE (LIGHT ORANGE SHADOW TOEARD ORANGE SQUARE) OF THE PRODUCT "SRS PROTOTYPE"-ELDERLY PEOPLE

C) Young disable man results- some words

With regard to the 23 year old man on wheelchair, it can be interesting to say that he is also amused by SRS system and possibilities but he also agree that to become a useful assistive device it still need a lot of development . However, the main difference with elderly people is that, despite the lowest Barthel index (33/100), he feels much more sure to be able to use the system by himself; despite he also has some handling and vision difficulties linked to his disability, he is in fact more skilled in using the UI-LOC device. Even if this is the point of view of a single man, this result is an encouraging, it make think that the SRS system could be exploited not only for elderly people but also for young people with motor disabilities.

4.2.2 PRIVATE CAREGIVERS RESULTS

The 12 private caregivers recruited were mostly the real family members of the elderly people recruited. They were the sons or sons in law in the majority of the cases, but also the husbands or wives, or in some cases aunts or uncles or grandsons. Only three private caregivers had not family relationship with the assisted. Table 12 and figure 26 report main features of private caregivers recruited.

TAB. 12 MAIN FEATURES OF ELDERLY PEOPLE RECRUITED

Main features of private caregivers recruited	
Age	Mean 54,1 (st. dev. = 13,4) Min 31, Max 66
Gender	6 males, 6 females
Education (0 null, 1 first, 2 low secondary, 3 high secondary, 4 degree)	Mean 3,5 (st. dev. = 0,5) Min 3, Max 4

FIG. 26 SOME OF THE PEOPLE RECRUITED USING THE INTERFACE UI-PRI DURING SCENARIOS EXECUTIONS.

A) Ad HoC Questionnaire results- private caregivers

Most of the caregivers were quite satisfied of the concept of the robot, considered valuable for the elderly to enhance the management of everyday tasks and support in case of scarce mobility. They also liked that the robot could give them the possibility to constantly monitor the health of the assisted person at a distance, thanks to visual control and speaking option, thus allowing for more freedom and more free time.

Nevertheless, the qualitative results highlighted that, though appreciating the idea in theory, there are some doubts linked to the limited range of activities accomplished by the robot at the moment, and to the fact of dealing with a prototype, that creates some worries concerning a potential failure of the robot in case of need. Indeed, though the majority of caregivers gave a positive feedback connected to a higher degree of security when living their assisted at home alone, qualitative results show they still have some fears.

The early stage of the robot development affected the caregivers' judgment on its usefulness: though this question scored 4.0, many interviewed commented on the scarce number of functions performed by the robot, stressing on the importance of its improvement and optimization. The same dissatisfaction regarded the time it took the robot to fulfill the assigned tasks, considered too long, especially taking into account an urgent or unexpected situation.

Evidently these considerations influenced also the hypothesis of effective usage and purchase of the robot: this time qualitative results are in line with quantitative results but help understand the reasons of disagreement. The people interviewed weren't really convinced of using or buying the robot, because its functions should be largely improved in order to cover all the needs of the old person, and many would rely on an external caregiver, instead or in addition to the robot, especially for emergencies and for night-time, as a person physically present at home is still supposed to better manage different situations (for example, in case of sudden need at night, the old person might call out for help, rather than looking for the electronic device).

As far as the usability of the interface is concerned, qualitative results integrate the score 4.0 of general satisfaction with it. Caregivers were basically satisfied with the interface, found it quite simple and intuitive to use, but during the interview emphasized that it wouldn't be that easy for the elderly, who are supposed not to be familiar with technology and less quick in learning.

Some of the interviewed also highlighted the complexity of the manual control of the robot's movements.

Figure 27 shows quantitative results (mean values and error bars) related to the questions addressed by 12 participants.

Table 13 reports the complete list of Usability problems emerged.

FIG. 27 QUANTITATIVE RESULTS OF AD HOC QUESTIONNAIRE (MEAN VALUES AND ERROR BARS) RELATED TO THE 12 PARTICIPANTS. SINGLE INDICATORS: GREEN IF SUCCESSFUL RESULT, RED IF NOT.

TAB. 13 USABILITY PROBLEMS EMERGED WITH UI-PRI

Usability problems of UI_PRI
Not all items were translated in Italian language, it is necessary to integrate the possibility to change language
Intuitive way to use manual navigation but not sufficient precision (e.g. moving the robot through a door). The remote control is difficult having only a 2D map displaying an icon of the robot.
• Skype not well integrated into the UI_PRI app (It is difficult to switch between the Skype and srs application in UI_PRI)
• UI_PRI video stream from the robot stopped working at several occasions. Restart was needed.
• UI_PRI – information about object on the tray is not provided
The status messages displayed on UI_PRI are confusing and not informative for the targeted user
Too poor info is provided on the apartment map,

B) Attrakdiff results - private caregivers

The following figures (28, 29, 30) report ATTRAKTIFF results obtained from private caregivers interviewed in relation to user interface UI-PRI. The last figure (figure 30) shows that the user interface UI-PRI was rated as "neutral". The classification here is not clearly "pragmatic" because the confidence interval overlaps into the neighboring character zone. The user is assisted by the product, however the value of pragmatic quality only reaches the average values. Consequently there is room for improvement in terms of usability. In terms of hedonic quality the character classification does clearly not apply because the confidence interval spills out over the character zone. The user is stimulated by this product, however the hedonic value is only average. Room for improvement exists in terms of hedonic quality. The confidence interval PQ is large. This could be attributed to limited sampling or to greatly differing product ratings. The users were less at one in their rating of pragmatic quality than in that of hedonic quality. Greatly differing ratings may be the result of participants having varying prior experience with the product itself (if possible) or with other similar products.

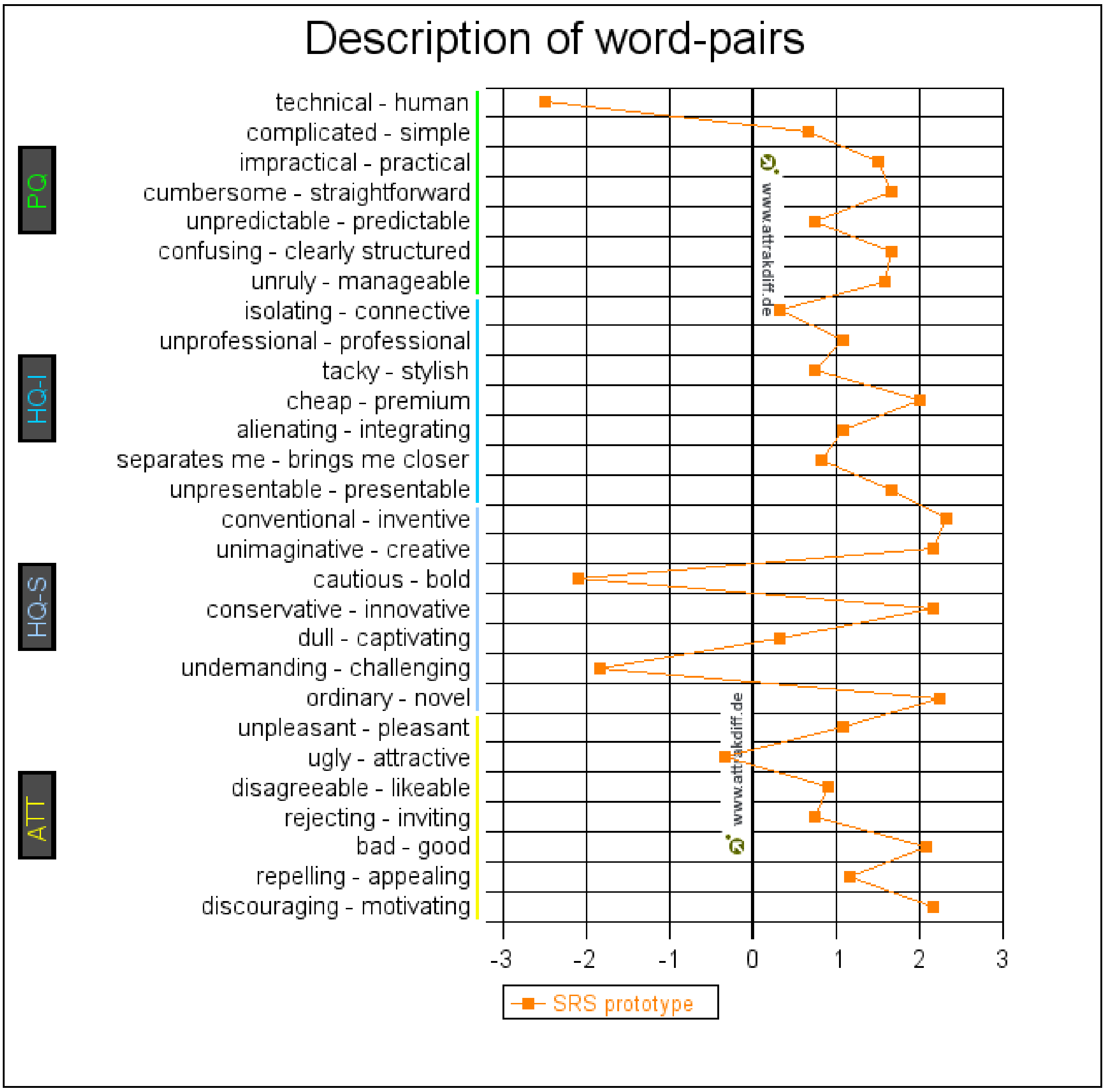


FIG. 28 ATTRAKDIFF RESULTS - DESCRIPTION OF WORD- PAIRS; MEAN VALUES - 12 PRIVATE CAREGIVERS

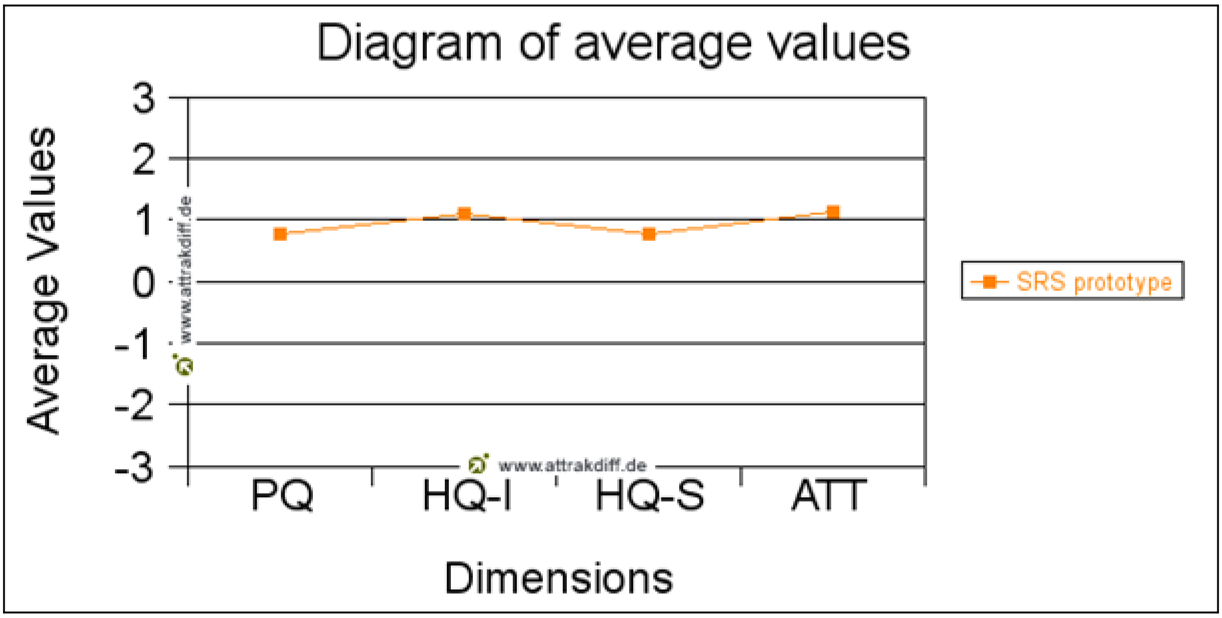


FIG. 29 DIAGRAM OF AVERAGE VALUES-PRIVATE CAREGIVERS

P Medium value of the dimensio with product "SRS"

Confidence rectangle

FIG. 30 PORTFOLIO WITH AVERAGE VALUES (ORANGE SQUARE REPORTING LETTER P) OF THE DIMENSIONS PQ AND HQ AND THE CONFIDENCE RECTANGLE (LIGHT ORANGE SHADOW TOEARD ORANGE SQUARE) OF THE PRODUCT "SRS PROTOTYPE"-PRIVATE CAREGIVERS

4.2.3 PROFESSIONAL OPERATORS RESULTS

Some Italian tele-assistance companies (e.g. Telbios http://www.telbios.com/ and Telesan http://www.telesan.it/) were contacted to participate in the study. They expressed interest related to the idea of such system but they were interested just to see a final product and not in participating in the test of a prototype at the beginning of its development.

However, to obtain useful feedback about current state of development of UI-PRO, five researchers (mean age 33,2, 1 female and 4 males) in the field of tele-medicine (biomedical/ electrical engineers) were recruited to test the prototype of the UI-PRO. They are experts on current tele-assistance devices and they work also on development of them (figure 31).

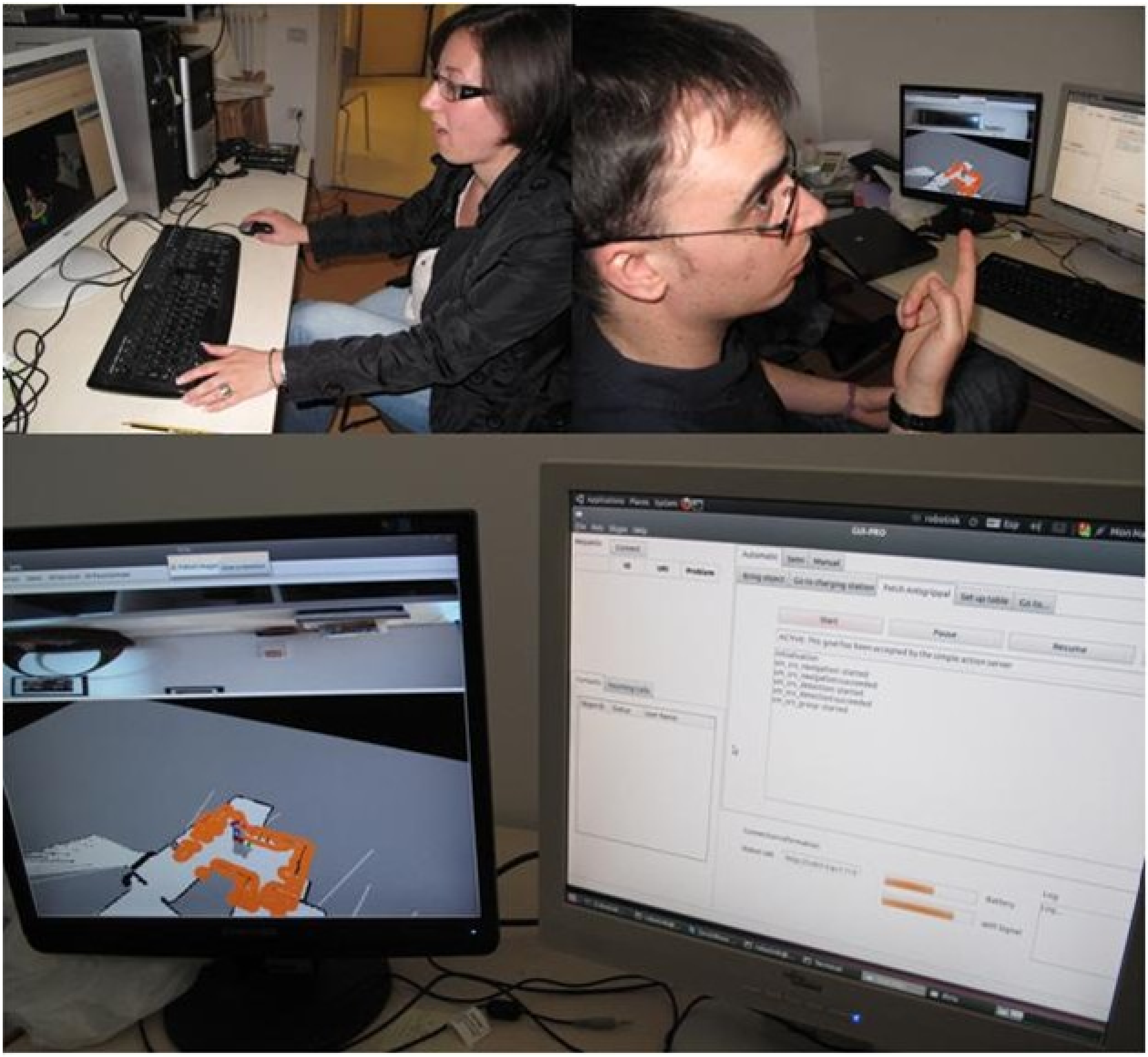


FIG. 31 TWO OF THE TELE.MEDICINE EXPERT INVOLVED IN UI-PRO TESTING

A) -Ad HoC Questionnaire results-professional operators

Ui-PRO interface is the interface that at the moment of Milan tests was at the lowest state of development. Main reasons for the bad results are related to usability of the Ui-PRO and are reported in table 14

TAB. 14 USABILITY PROBLEMS EMERGED WITH UI-PRI

Usability problems of UI_PRI
Datas are not shown in real time (or a least in useful time)
There are many communication problems (Skype drops out, contact with the robot drops out)
Interface is not user friendly even for engineers, it is difficult to be understood without a long training.
Bandwidth problems do not allow the synchronous use of video and virtual reality considered mandatory
It's not possible to see details of the robotic arm and to properly control it (avoiding hurts for example)
It is not possible to manually control the robot, only autonomous mode is working at the moment.
The map visualization is too poor, more details of the house are needed.
Virtual reality must be improved; 3d vision is not complete (in the map you see the robot 3D, while map is 2D, so there are no references related to height of furniture)
It would be better if the elderly could talk directly to the robot and not only through the robot interface (see already existing Giraffe robot)
There is no way to teach the robot new objects

Apart from the user interface, the people interviewed in general liked the concept of SRS, the idea of the possibility of audio-video feedback to check consciousness status of the assisted together with the possibility to execute some tasks as reaching objects located in difficult places to be reached.

However they were not so enthusiastic as the other two groups (elderly people and private caregivers); they are more aware of current state of development of other comparable products and projects, and so they are not so easily impressionable just from novelty represented by SRS (see Giraffe robot, Smart home facilities, different kind of health status monitoring systems…).

They judged the current status of development as not sufficient, being the robot not effective in most of the cases even in performing just one and very simple task. The SRS solution therefore was not considered useful at this stage. They judged the "*Technology development still far from ensuring reasonable time execution, accuracy, stability and execution modality"*

In particular even when the robot succeeded in completing the task, as the other groups of participants they were not satisfied at all with the time for compete the task *"To go and bring a medicine it took 7 minutes, to be useful it should take maximum 2 minutes, otherwise time become comparable with ambulance coming (10-15 minutes)".*

The robot was also judged too big, and because using an industrial arm not suitable at all for houses; moreover It's behavior was considered unpredictable *"The robot dances, it's not clear it's direction, it's path".*

Figure 32 shows quantitative results (mean values and error bars) related to the questions addressed by 5 participants confirming the qualitative expressed view.

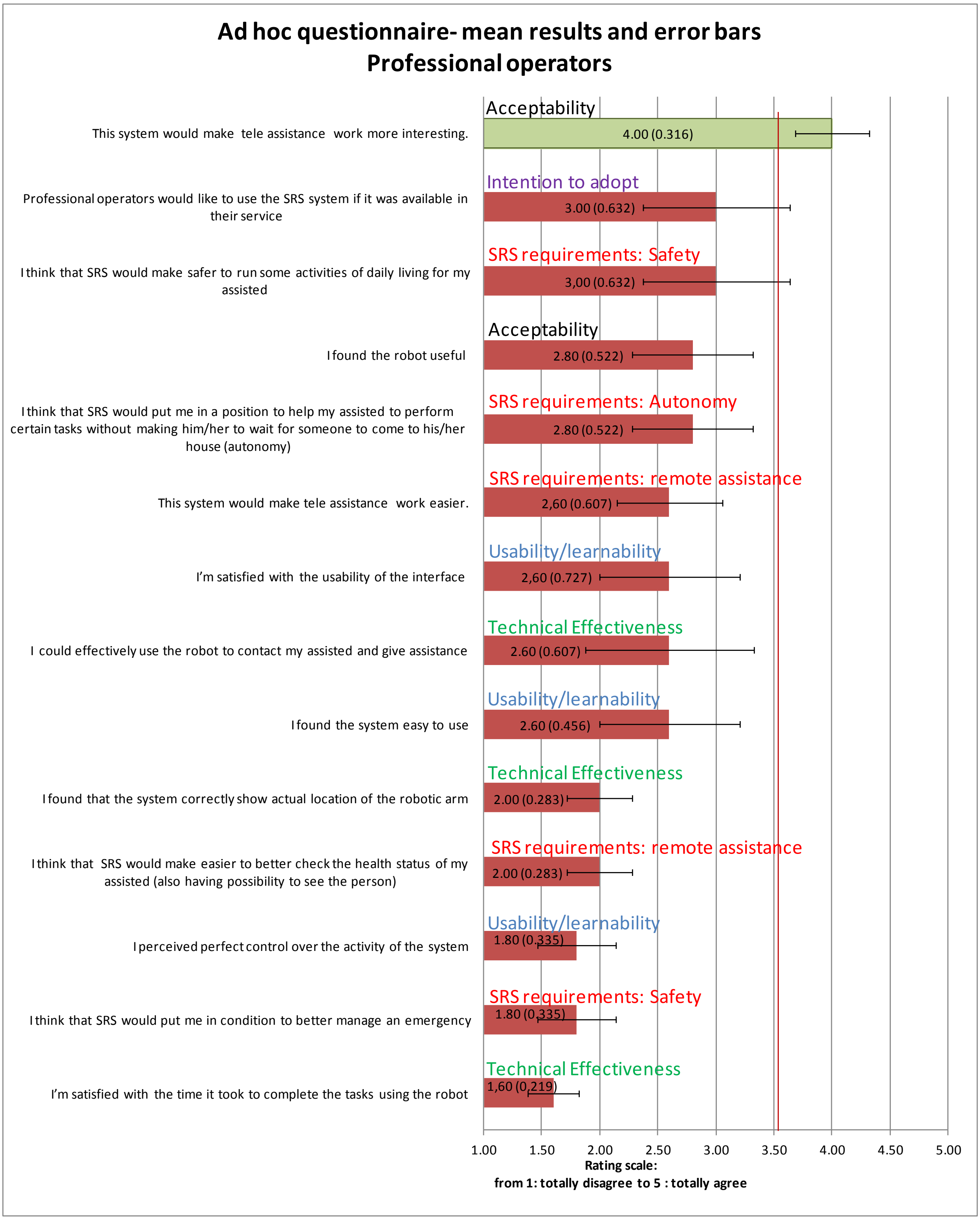


FIG. 32 QUANTITATIVE RESULTS OF AD HOC QUESTIONNAIRE (MEAN VALUES AND ERROR BARS) RELATED TO THE 5 PARTICIPANTS. SINGLE INDICATORS: GREEN IF SUCCESSFUL RESULT, RED IF NOT.

Attrakdiff results- professional remote operators

The following figures (33, 34, 35) report ATTRAKTIFF results obtained from professional operators interviewed in relation to user interface UI-PRO. The last figure (figure 35) shows that the user interface UI-PRO was rated as "neutral". The classification here is not clearly "pragmatic" because the confidence interval overlaps into the neighboring character zone. The user is assisted by the product, however the value of pragmatic quality only reaches the average values. Consequently there is definite room for improvement. In terms of hedonic quality the character classification does clearly not apply because the confidence interval spills out over the character zone. The user is stimulated by this product, however the hedonic value is only average. Room for improvement exists in terms of hedonic quality. The confidence intervals of both dimensions are large. This could be attributed to limited sampling or to greatly differing product ratings. The users are at one in their ratings of both dimensions.

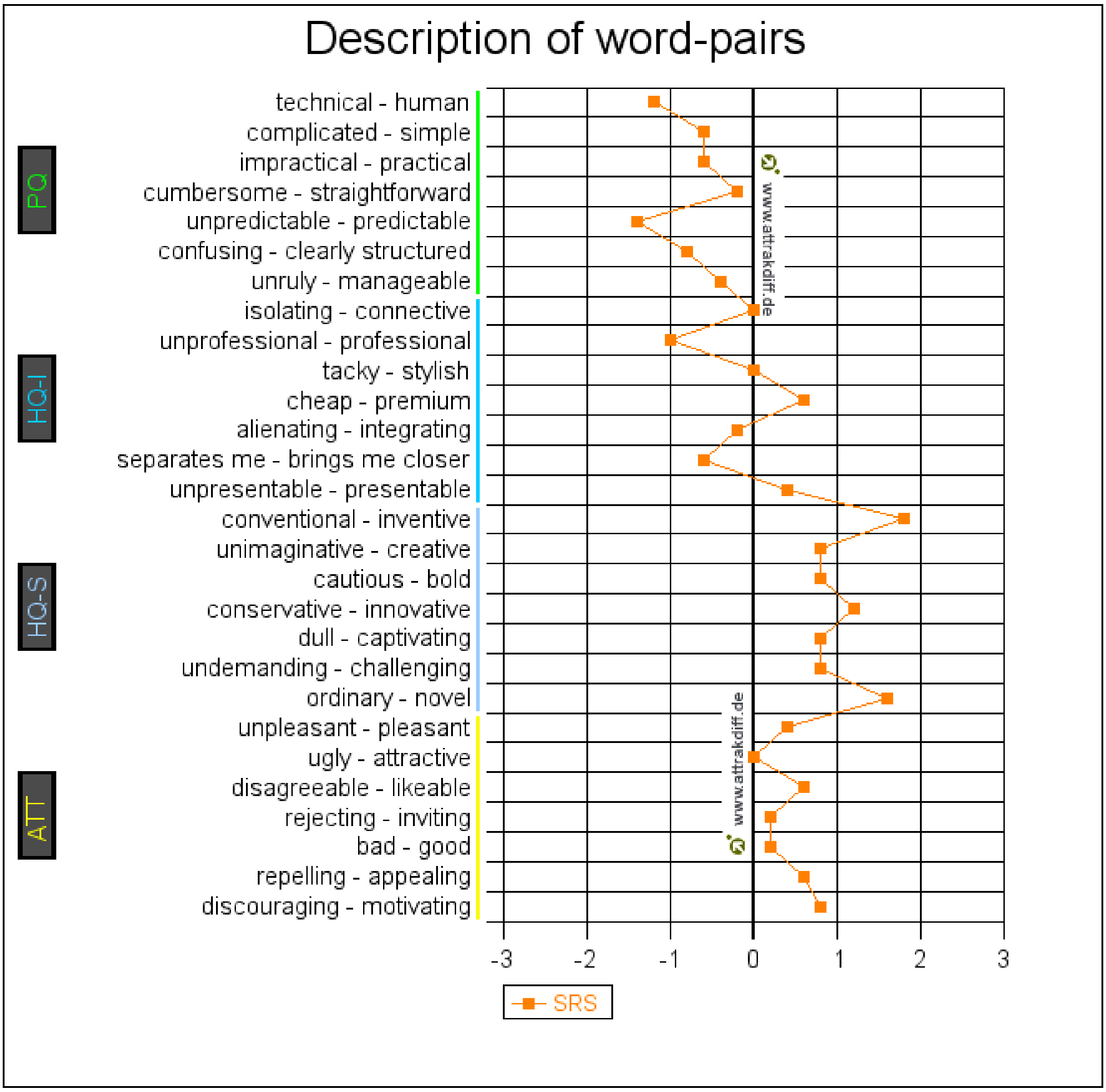


FIG. 33 ATTRAKDIFF RESULTS - DESCRIPTION OF WORD- PAIRS; MEAN VALUES – 5 PROFESSIONAL OPERATORS

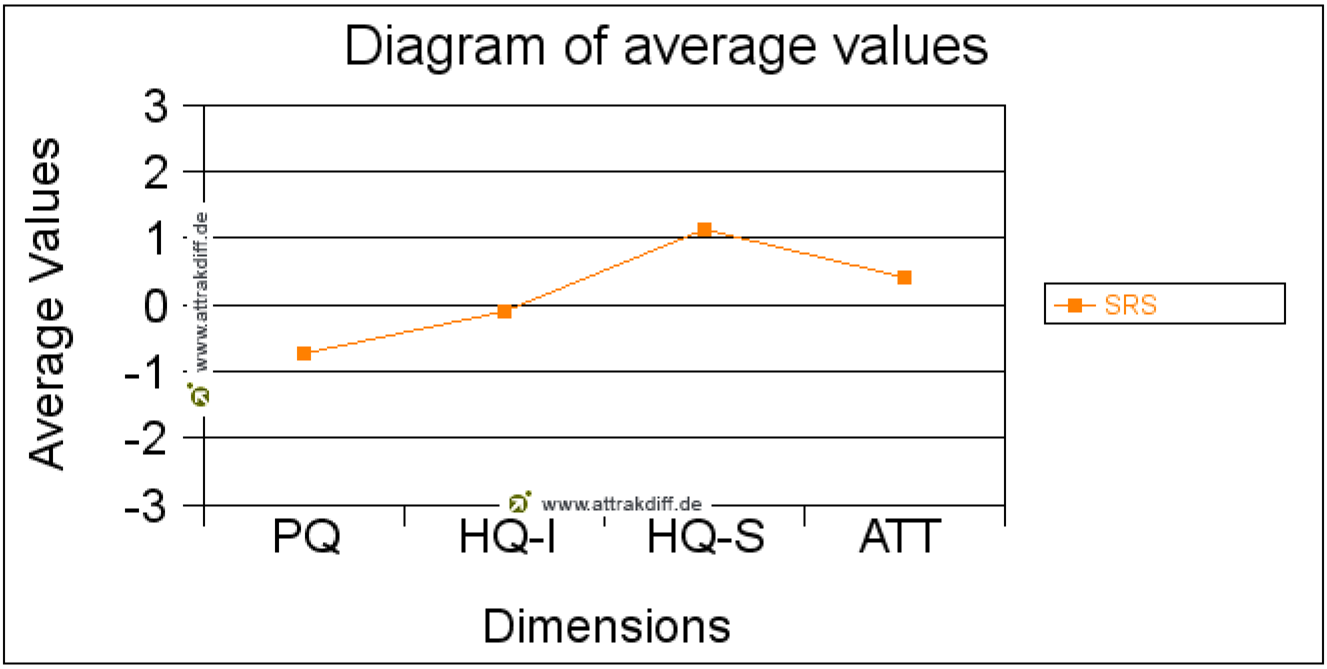


FIG. 34 DIAGRAM OF AVERAGE VALUES-PROFESSIONAL OPERATORS

FIG. 35 PORTFOLIO WITH AVERAGE VALUES (ORANGE SQUARE REPORTING LETTER P) OF THE DIMENSIONS PQ AND HQ AND THE CONFIDENCE RECTANGLE (LIGHT ORANGE SHADOW TOEARD ORANGE SQUARE) OF THE PRODUCT "SRS PROTOTYPE"-PROFESSIONAL OPERATORS

4.3 TECHNICAL ISSUES

As stated in section 4.1 of the document, tests in Milan were focusing on user experience and robot acceptance. Even if the focus was on user experience, tests provided technical feedbacks, especially about the integration and performance of all available components. Their fundamental data will ensure the right direction of future development and integration effort.

As a result of the thorough testing of SRS system with many users, as reported within this deliverable, a number of technical problems were observed which, to varying degrees, hindered the successful execution of the test scenario and necessitated involvement of the technical personnel to correct the problem.

Subsequent analysis of the tests and the associated problems showed that they can be classified in the following groups, listed according to their effect on the execution of the tests:

- Usability issues of the interfaces – moderate impact
- Reliability issues – moderate impact
- Efficiency issue of the overall system – minor impact

Tests performed with users during the first weeks of May, showed technical issues due to the challenge of bringing the robot in a new unknown environment and relying on many prototype-level components developed by different partners. Some of these issues had relevant impacts on users' experience while others did not affect the results at all. The purpose of this analysis is to show in which amount the issues affected the tests.

During the five test days, the three version of "fetch and bring medicine" scenario have been run over and over with different users to establish the level of impact and acceptance. Each version of the scenario is composed by subtask performed either by users or by robot.

For more information the Tab. 15 below provides a description of each task for each version of the scenario.

TAB. 15: SCENARIOS'SUBTASKS

Scenario: "autonomous fetch and bring of the medicine from the shelf to the sofa"	
Version 1: **Elderly improving their Autonomy at home**	Elderly user selects carry task on local UI
	Elderly user sends command to Robot
	Robot navigates correctly to destination
	Robot tries to identify the object's position
	Robot grasps the object
	Robot places the object on the tray
	Robot navigates to elderly user position
Version 2: **Private caregivers monitoring situation and remotely assisting**	Relative calls the elderly with UI-PRI
	Relative send command to robot with UI-PRI
	Robot navigates correctly to destination
	Robot tries to identify the object's position
	Robot grasps the object
	Robot places the object on the tray
	Robot navigates to elderly user position
Version 3: **Professional operators**	Elderly calls remote assistance
	Remote operator sends command to robot
	Robot navigates correctly to destination

managing emergency	Remote operator sends fetch medicine command to robot
	Robot tries to identify the object's position
	Robot grasps the objects
	Robot places the object on the tray
	Robot navigates to elderly user position

Of 19 tests performed, four were executed without any issues in the subtasks, as shown in Figure 36, while the remaining 15 had issues on different levels which could affected the test outcome. In some cases, e.g. when dealing navigation issues, the tests were still resumed with all successive subtasks, but if the failure was major (such as the robot not being able to identify the object to be picked) the tests were not completed and it was not possible to test all successive sub tasks. Some tests report more than one issue when more tasks failed due to problems (for instance grasping and navigation). This is due to the complexity of the scenario.

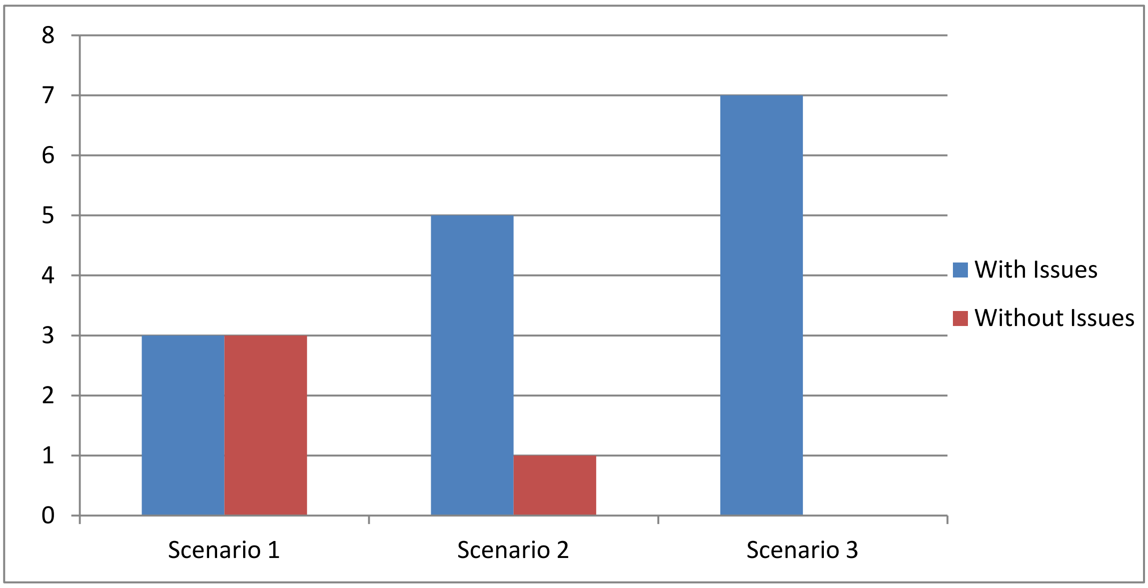


FIG. 36 FAILURE COUNT FOR EACH SCENARIO VERSION

Figure 37 shows how the three major failures, namely navigation grasping and network connection problems, affected each of the three scenarios. Scenario 3 was the only one affected by network connection errors, mainly related to the PRO-UI. The total number of issues is grater then the total test executed because in some case more issues occurred during the same scenario (for example when robot failed to grasp object and later to navigate to elderly position).

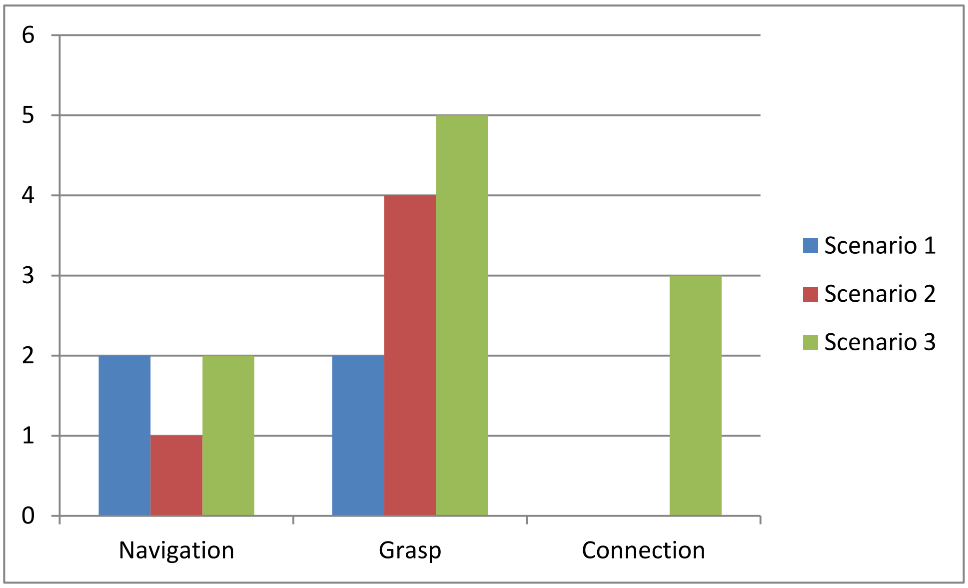


FIG. 37 ISSUES COUNT

Figure 38 shows the incidence of each kind of failure over the total issues. It can be seen that most of the problems were related to the failure of the grasping subtask. Even if there was a high percentage of grasping or navigation failures in the tests, it should be considered that most of the time, to complete each scenario and allow users to give a correct feedback, the local human operator intervened to put the object on the tray if the grasp was unsuccessful or to manually correct the robot navigation especially when passing through the door.

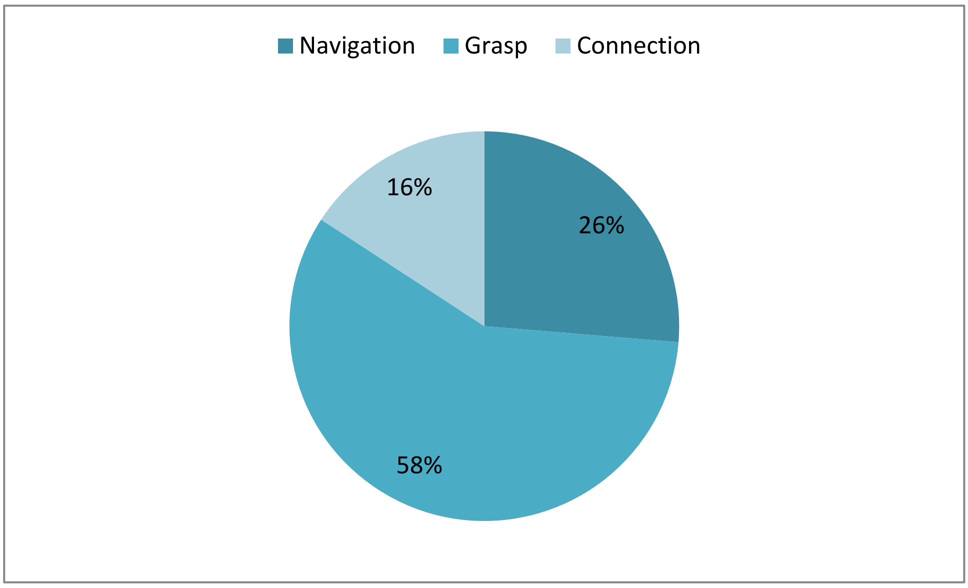


FIG. 38 PERCENTAGE DISTRIBUTION OF THE ISSUES OVER THE TOTAL

What is not shown in the graph is the users' feedback about the performance of the whole system. Performance issues were noted by many users who perceived grasping and navigation as being too slow. In conclusion the following issues were identified:

- User interaction issues: Robot not correctly receiving tasks from users interface.
- Navigation issues: these issues are related to robot movement to a specific destination such as the object location or the user location.
- Fetch and carry issues: tasks performed once the robot is near the location of the object to be picked up and for some reason robot is not able either to identify, to grasp or to place the object in the tray.
- Performance issues: behind the count of failed tasks there is a feedback related to the performance of robot. Overall the feedback from users noted that robot movement appeared too slow and the time required to fetch and carry the medicine was high.

The issues impacting the results of the tests in Milan are summarized by the cause diagram shown in Figure 39 and can be commented as above.

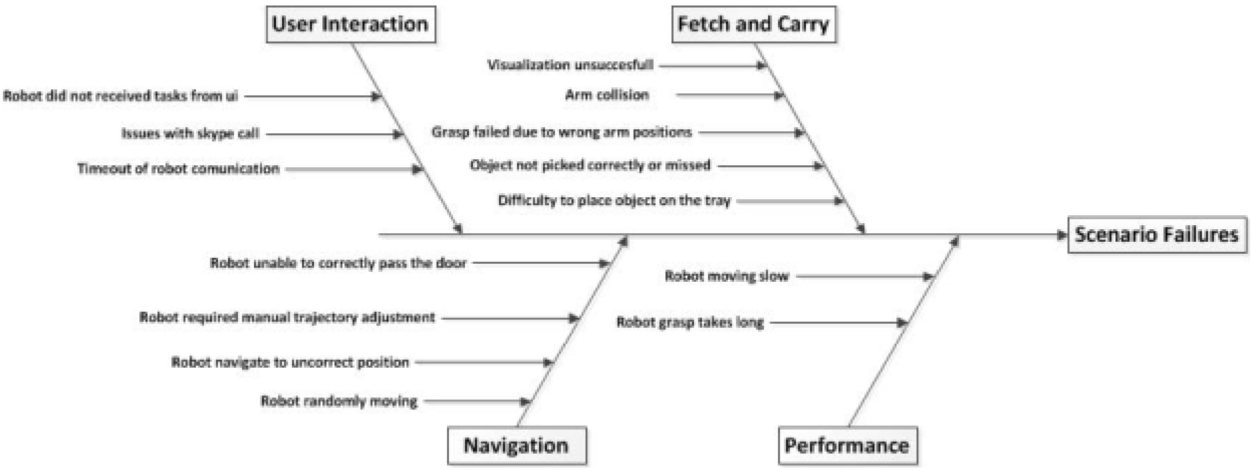


FIG. 39 FAILURE ANALYSIS

4.4 DISCUSSION

The system is perceived interesting and potentially very useful by all user groups. Acceptance is high and this can be highlighted as the better result obtained in this first set of tests with the integrated prototype. Elderly people and private caregivers are more enthusiastic about the novelty represented by SRS system with respect to professional operators; people belonging to this last group are more aware of current state of development of other comparable products and projects, and so they are not so easily impressionable just from novelty represented by SRS.

With regard to the 23 year old man on wheelchair, also assessed as potential local user, it can be interesting to say that he is also amused by SRS system future possibilities, and the main difference found with regard to elderly people is that he feels much more sure to be able to use the system by himself, not fearing the use of the UI-LOC interface; so it could be very interesting to explore if exploitation of SRS system could be possible also for young people with motor disabilities.

Local users (elderly people and young man) and private caregivers were made aware that the presented robot was, at this stage, only a prototype, and so not fully completed and ready for market. Therefore, it could be interpreted that their judgment of the SRS concept was not as positive as it may first seen for usability, usefulness and effectiveness. as their feedback may have compensated for or taken into consideration the prototype status of the robot. This is also reflected from the low score obtained by indicator intention to adopt, that didn’t reach success threshold of 3.5. Users negative judgment respect the current state of development, are part due to usability of interfaces, as well as the technical problems that, at times and to varying degrees, hindered the successful execution of the test scenario and necessitated involvement of the technical personnel to correct the problem.

Subsequent analysis of the tests and the associated problems showed that they can be classified in the following groups, listed according to their effect on the execution of the tests:

- Usability issues of the interfaces – moderate impact
- Reliability issues – moderate impact
- Efficiency issue of the overall system – minor impact

Summarizing, these groups contain the following problems:

USABILITY ISSUES OF THE SRS INTERFACE DEVICES (UI_LOC, UI_PRI, UI_PRO)

A number of problems were identified when the users were controlling the robot or communicating through the various interfaces. These ranged from small problems, e.g. unfamiliarity with the interface and nuisance software bugs resulting in error messages to very serious problems that disabled part of the functionality, e.g. UI_PRO crashing when triggering certain actions on the robot. In addition, the more elderly users didn't cope well with the touch screen buttons on the UI_LOC device. The usability of the system for more elderly users, who are most likely to be the end-users of the SRS system, is there an area of particular concern. The problems have been discussed by the consortium partners and appropriate measures are planned to eliminate these problems in the future.

RELIABILITY ISSUES OF THE SRS SYSTEM

During the user tests the SRS system suffered from a number of intermittent problems related to the robot's task performance. On the 70 % of occasions the robot was unable to complete its task successfully without any issue. As this involves a number of software and hardware modules and because of the random and intermittent occurrences it was impossible to isolate the cause(s) on-site. Therefore further rigorous testing and investigation is needed to identify the exact source. The robustness of the system under different conditions is considered to be of high importance for the successful application and exploitation. Therefore technical partners will need to carry out further testing and corrective actions in to ensure its reliability before any further tests with users are carried out.

EFFICIENCY ISSUES OF THE SRS SYSTEM

As user feedback has shown in this deliverable users did express some dissatisfaction with the overall efficiency of the system. Whilst often the task would be completed within a short period of time there were occasions when the task could take 10 minutes to complete (or when there was a technical problem not complete at all). Therefore testing and investigation is needed to identify and eliminated the causes of the problems to ensure the robustness efficiency of the system.

In conclusion, using the successful indicators specified in D6.1, 11 out 18 targets was met for the elderly people, 12 out 14 targets was met for private caregivers, only 1 out of 14 targets was met for professional operators. On the other hand, the technical partners in the consortium are creating an action plan to address and solve the identified problems in timely fashion before the next consortium meetings. By investigating and addressing all the issues, it will enable improved user tests to achieve all successful indicators.

An investment of forces moreover in technical development was needed and so other two steps of testing iteration have then planned focusing on UI-PRO, which is the interface that at the time of Milan tests obtained lower results of acceptance by the users.

5 SECOND VISUALIZATION TRIALS

5.1 METHOD AND OBJECTIVES

The aim of the second visualization trials was to test the updated UI-PRO interface. After the first trials (par. 2), several usability suggestions were made. The interface was adapted according to the suggestions in close contact between user partners and technical partners.

In contrast with the old web based interface, the updated UI_PRO was based on a window environment. Developed in C++ language, it offered more possibilities to implement new utilities and tools. Some of the main improvements of the new version were:

- Communication between the tele-operator and the assisted user through the Skype API;
- Autonomous commands such as "move to kitchen", "grasp object", "search object", etc.;
- More user friendly appearance;
- Usability improvements;
- More information from the robot (feedback) visible for the tele-operator;
- ROS Rviz module integrated as 3D viewer (more complex and dynamic that the old web viewer);

The Rviz module was ready to add new displays through using plug-ins. These plug-ins offered important tools like a joystick to move the robot, or some assisted arm manipulation tools to facilitate the grasping of objects. Moreover it provided powerful 3D tools which offered much better information about the environment surrounding the robot: point clouds, 3D objects, shadows, distance indicators between the robot and some physical objects, etc.

Summarizing, the new interface was a new and absolutely different concept that the old one, both in operation and capabilities terms, as well as in its user friendly appearance.

The second UI-PRO was released to be tested, after having passed some internal functional tests which assured the correct functionality of the system, previously to be tested by the user.

Visualization tests in San Sebastian were planned to address the main peculiarities of the interface with real users (professional remote operators and usability experts), and to study the effectiveness, usability and user acceptance of the robot manipulation interface.

The experimental objectives of the tests are the same as in the previous trials:

- Determine the accuracy and efficiency of the professional manipulator interface in performing tasks (selected from scenarios).

- Evaluate the feedback provided to the professional operators through the visualization system of the UI-PRO device during tasks executions.

- Collect data about users' subjective perception on the robotic arm and on tasks based on arm manipulation.

- Compare the first prototype results to the second prototype results to evaluate the improvement of the system.

5.1.1 Protocol

To achieve the objectives in this trial, the same set up as in the first trial was carried out. Tests with advanced prototypes are based on an experimental protocol involving 13 participants (recruitment criteria are defined into D6.1-1), the same as possible that participated in the first trial. They will compose the remote operators (RO-PRO) sample: potential professional users of the SRS system. The same experimental protocol as in the first trial was carried out (for a more specific description, DELIVERABLE D6.1b Testing site preparation and protocol development).

The trials focused on grasping and fetching things, as well as on the visualization systems. The main outcome variables to measure were:

- Acceptance: outcome expectations, advantages/disadvantages perceived, attractiveness, comfort perception, eligibility (intention to adopt) and perceived usefulness.
- Usability: performance evaluation, the success on task execution will be recorded, as well as the time required for its completion, and the errors and/or usability difficulties that may occur within the process.
- Acceptability: Emotional experience of interaction with the system.

The tests were carried out during month 36 - December 2012: the tests were conducted with potential professional operators by using the UI_PRO for robot visualization and grasping. In this test, the robotic arm was not directly needed, it was simulated on a laptop. The usage of UI_PRO was investigated in order to determine if it could provide the professional users a tool to control the SRS system. One of the main features of UI_PRO is in fact its capability to provide a "visualization of environment". This visualization was perceived not only through the cameras of the COB, but also a visualization of the simulated environment by means of the remaining sensors was captured.

Although these trials mainly focused on the manipulation abilities of the system, they gave also the chance to test the human-interface devices developed so far, concentrating on the specific context of the manipulation tasks.

Each test session was supervised by at least two SRS Operators, one from Robotnik taking care of technological aspects, and one from Ingema coordinating the experimental procedure and registering qualitative data.

5.1.2 Evaluation procedures

The main parameters to be measured in these trials were:

- Usability (Efficacy and efficiency)
- Acceptance (Comfort perception)
- Acceptability (from a Psychological/emotional point of view)

VISUALIZATION TESTS Variables:

Demographics (age, education, occupation, housing situation)

- Questions based on SoTU subscale: Experience with technology
- Questions based on SoTU subscale: Demonstration
- Observation
- Notes

Usability (Efficacy and efficiency)

- Time
- Task Completion
- Errors
- Questions and hints

Acceptability

- Attrakdiff User Experience questionnaire

Acceptance

- Intention to use
- Usability perception
- Accessibility perception
- Intention to use

Usability

All the measures were completed with information collected by Ingema staff through different observation, interview and assessment by specific questionnaires detailed at D6.1.

5.1.3 Participants

We have collected data from professionals working in areas similar to those involved in SRS. In concrete, we recruited 13 professionals. 8 professionals worked for a tele-assistance company (SaludNova, http://www.saludnova.com/) and 5 for Matia - Ingema professionals. 7 woman and 6 men participated, all of them were medium-aged adults (mean = 30.85 years; st. dev. = 4.79). All of them have high educative level (mean =23.30 years of education; st. dev. = 3.97) and have a job with responsibilities from medium to high.

All of them have university graduates or equivalent (mean =23.30 years of education; st. dev. = 3.97) and have a job with medium to high responsibilities related to tele-care and /or gerontology. All they reported to have jobs with high cognitive and social requirements, and low physical requirements.

Their experience with technology is high. They use all the devices included in the questionnaire (washing machine, dish washer, vacuum cleaner, TV, mobile phone / smart phone, PC / laptop), and they report to have been using PC / laptop a mean of 16.77 years (st. dev. = 6,27; min. 8 years, max. 30 years).

The users were the same ones, as much as possible, whom took part in the first study to perform the usability tests, reported in paragraph 2.2.

5.2 USABILITY, ACCESSIBILITY AND ACCEPTABILITY ASSESSMENT

In the second visualization trials we have followed the same procedure as in the first phase visualization trials in order to be able to compare the results between the first prototype and the second one. This quasi-experimental design would facilitate to track the usability changes between the first interface version and the second one.

Three tasks were established in order to let the participants use the system in the same settings as the first time aiming to simulate the natural usage scenarios.

Employees from SaludNova were interviewed in the facilities of the company. The rest of the participants were assessed in the INGEMA premises.

5.2.1 INSTRUMENTS AND MEASUREMENTS

In these trials, we used the same measurement instruments that we used in the first trials: AttrakDiff and structured ad-hoc questions. Apart from the ad hoc post-test questionnaires, an observational procedure was developed to take into account primary user-system interaction characteristics. Usability of the system was evaluated by assessing the users' tasks performance regarding: task completion, errors, time, hints needed, comments, and usability difficulties observed.

5.2.2 USABILITY AND ACCESSIBILITY RESULTS

Three specific scenarios were developed based on the general scenarios of the project, in order to allow the user interact with the system: grasping, leaving and moving scenario. Complementarily, ad-hoc quantitative questions were done about these issues after the three scenarios.

A) Grasping scenario

TAB. 16 USER TASK PERFORMANCES.

Task Completion	13 out of 13 completed the task.
Time	Mean= 325.85 seconds st. dev. = 129.57 seconds. The fastest of the users

	completed the task in 210 seconds meanwhile the slowest of the participants that completed the task needed 600 seconds to complete the task.
Errors	No errors were made by the users.
Questions asked	0.30 questions on average were made by the users. That means that nearly one out of 3 users asked at least a question about the functionality or how to perform the task after having been explained.
Hints given	No hints were needed to perform this task.
Facial expression:	The expression of the participants was neutral o positive in most of the cases. Just two participants seemed to be frustrated about using the system.

Comments (made by the user or the observer) and difficulties:

Verbal transcriptions.

- *"The camera movement is weird. It doesn't reach the object".*
- *"Is the robot doing anything or has got blocked?".*
- *"Should be a notification about the process status".*
- *"It's easy to manipulate, the interface is functional and correct, but the task takes a lot of time to be finished and finally it doesn't grab the object".*
- *"It could be a grabbing notification in a "loading bar" way, something that could give feedback about the process is still working and that didn't get blocked".*
- *"Interface is intuitive but not the Gazebo vision".*
- *"The interface is not intuitive enough, I could not finish the tasks without following the instructions".*
- *"Is grabbing procedure already started?".*
- *"If this is the real speed of the robot it will not be used by anyone".*
- *"If the task is somehow urgent, the robot will not accomplish at this pace".*
- *"How the real life objects will be distinguished? How can we distinguish between a milk tetrabrik and a wine one?".*

- *"I like the process list of process".*
- *"The user should be allowed to place an axis starting point from which start the camera movement".*

Difficulties while performing the task.

- The user suggests improving the camera movement.
- In several occasions, the automatic grabbing control pull down menu it's automatically activated and some users suggested that the tabs should be clicked in order to be displayed.
- In some occasion, the robot stated that the item was not grabbed when it actually was.
- The Gazebo camera control is complicated and not intuitive.
- The users do not know when the process starts and if the process still works or has got stuck.
- Messages should give more detail of the ongoing process.
- The task performance is slow and it is frustrating for some participants.
- They are concerned about the privacy if there is the need to have cameras at home.
- The users are concerned about the real life objects appearance and how the robot could distinguish between objects.
- One user suggested having some labels with the name of the object, and other labels named: where.

Summarizing: the most concerning difficulties and the most repeated ones are: slowness of the robot performing the task, and the need for more feedback on the process.

B) Leaving Scenario:

TAB. 17 USER TASK PERFORMANCES.

Task Completion	13 out of 13 completed the task.
Time	Mean= 302.53 seconds st. dev. = 93.22 seconds. The fastest of the users completed the task in 186 seconds meanwhile the slowest of the participants that completed the task needed 513 seconds to complete the task. This was the time needed by the user and the system to perform the whole task. On the other hand, just M=3.15 seconds were needed by the users to perform their part.
Errors	0.30 errors on average were made by the users.

Questions asked	0.46 questions on average were made by the users. That means that nearly one in two users ask at least a question about the functionality or how to perform the task after having been explained.
Hints given	0.38 hints were given on average. Nearly a third needed more hints to complete the task.
Facial expression	The expression of the participants was neutral o positive in most of the cases. Just two participants seemed to be frustrated about using the system.

Comments (made by the user or the observer) and difficulties:

Verbal transcriptions

- *"The display menu is very complicated and tedious to be followed. The ideal thing should be to click in some point of the map and the robot went there".*
- *"Rviz screen seems to be unfinished".*
- *"To have the mouse clicked (as in drag and drop) to move the robot is not comfortable".*
- *"Where are the tools?".*
- *"I prefer to use the yellow circle, is much more precise".*
- *"I don't know the distances, there should be a notification about the collisions".*
- *"There are a lot of options to perform a single movement".*
- *"The interface slows everything and makes everything more difficult to learn".*
- *"At first is kind of messy, but in the future it's supposed to be manipulated in more intuitive".*
- *"The display interface is not much intuitive".*
- *"I don't understand the display instructions".*
- *"Is the robot already moving?".*
- *"It has a bad design, I don't like it. The displays window is hard to understand and interact. I would suggest just one window for everything and changing everything in tabs. This way is too slow".*
- *"I don't understand why should I have to have the button clicked, why not just click and the robot moves in this direction?"*

- *"The steps are easy to follow, is the robot which is slow".*

Difficulties while performing the task.

- The user did not understand the interactive markers instructions and need more detailed explanations.
- Some participants cannot see the kitchen counter in the Rviz.
- Rviz y gazebo should be integrated.
- Arrows did not appear for the hand manipulation.
- The steps seemed to be too much long and not intuitive.
- The cameras did not see the obstacles making easy to collide.
- Some participants have problems to understand the axis and how does the robot movement works.
- The participants had difficulties to perform the robot movement.
- Gazebo and Rviz should be in the same window.
- We should count on more reliable and precise sensors and cameras.
- The drag and drop kind of movement is not so easy as a point and click.

Functional Problems:

- A failure appeared as a blue square in the Rviz.
- The Rviz is empty.
- A user clicked in the red button instead of launch it. Closes gazebo by error.

Summarizing

The Robot performance/movement speed is slow for the user manipulation. The user has to wait until the system performs the whole task taking an unacceptable amount of time. Probably if the user only gives the instructions and the system performs the maneuver, the time the system needs to perform the task would not be a problem, but in this case the user has to move the robot through the whole movement sequence. This could be boring, frustrating and too much attention demanding to be performed with safety and reliability.

The 3d representation could be improved by having it only in one window.

Participants need more instructions than the ones of the program.

C) Moving scenario:

TAB. 18 USER TASK PERFORMANCES.

Task Completion	11 out of 13 completed the task.
Time	Mean= 202.76 seconds st. dev. =33.9 seconds. The fastest of the users completed the task in 100 seconds meanwhile the slowest of the participants that completed the task needed 480 seconds to complete the task.
Errors	Just one user fail in one occasion.
Questions asked	0.46 questions on average were made by the users. That means that nearly one in two users ask at least a question about the functionality or how to perform the task after having been explained.
Hints given	1.38 hints were given on average. Nearly a third more hints that the users asked for.
Facial expression	5 out of 13 users showed frustrated feelings when performing the task

Comments (made by the user or the observer) and difficulties:

Verbal transcriptions are reported below:

- *“Where is the table, it cannot be seen”.*
- *“Robot is too slow”.*
- *“The camera cannot be seen well”.*
- *“The best could be not to have the button pressed to move the robot”.*
- *“The turning direction in the blue circle changes when 180º are overtaken”.*
- *“Navigation is difficult”.*
- *“It should be possible to point the place where we want the robot to go and just point it”.*
- *“Gazebo and Rviz should be in the same screen, it has no logic to have to be looking both screens at the same time”.*
- *“The ideal should be not allowing robot movements if there could be collision involved”.*

- *“The program is not excessively slow, but should be faster and possible to point the direction instead of having to move it there”.*
- *“The interface should be cleaned up and just showing what the user needs to interact”.*
- *“Why I cannot see the whole room in the DRivz”.*
- *“The robot should notify the user when it hits anything, not just stopping”.*
- *“Red and blue colors are not a good option for accessibility since they can be confused by people with color impairment”.*
- *“Gazebo and Rviz should be in the same window”.*

Difficulties while performing the task.

- The users point out that the system should detect the objects and notify its presence not only stopping before hitting.
- Some participants have problems keeping the button pressed and stop moving the robot by releasing it.
- Two screens are not very usable, one could be a better solution.
- The interface is difficult to use without formation.
- The interface should be in the language of the country.
- Distances were difficult to perceive.
- One user become frustrated by the pace of the robot and stopped using it.
- The users needed to switch between windows, and that complicated the robot manipulation.
- Even when the robot has sensors to find obstacles they should be seen by the user.
- There should be buttons with predetermined turning options (45º, 90º, 180º)
- In the Rviz the room should be seen.

5.2.3 Acceptability results

All the participants answered the ATRAKKDIFF instrument and the questions developed ad hoc to assess acceptability. The results can be observed in the following tables.

TAB. 19 ATRAKKDIFF INSTRUMENT RESULTS.

	1	2	3	4	5	6	7	
annoying	3	2	5	2		1		**enjoyable**
not understandable	-	2	4	3	1	2	1	**understandable**
creative	1	2	1	5	3	1	-	**dull**
easy to learn	2	5	2		2	2	-	**difficult to learn**
valuable	-	3	2	3	3	2	-	**inferior**
boring	6	-	2	2	3	-	-	**exiting**
not interesting	1	2	2	3	3	1	1	**interesting**
unpredictable	2	2	2	2	3	1	1	**predictable**
fast	-	-	-	-	-	6	10	**slow**
inventive	2	3	4	4	-	-	-	**conventional**
obstructive	1	2	-	4	3	2	-	**supportive**
good	1	1	5	1	2	3	-	**bad**
complicated		1	1	4	2	1	1	**easy**
unlikable	1	-	4	4	4	-	-	**pleasing**
usual	-	1	-	4	3	3	2	**leading edge**
pleasant	2	1	2	5	2	1	-	**unpleasant**
secure	-	2	2	4	2	2	1	**not secure**
motivating	-	2	2	3	3	2	1	**demotivating**
meets expectations	-	1	2	3	1	4	2	**does not meet expects.**

efficient	-	-	1	5	2	2	3	**inefficient**
clear	-	-	5	1	-	5	2	**confusing**
impractical	1	1	2	1	7	1	-	**practical**
organized	1	2	-	2	4	1	3	**cluttered**
attractive	2	4	2	2	2	-	-	**unattractive**
friendly	-	-	2	6	3	1	1	**unfriendly**
conservative	-	-	-	2	2	7	2	**innovative**

Responses in the AttrakDiff questionnaire were found to be clustered around the mean values. Professionals agree that the technology is innovative, attractive and interesting, but also slow, and probably due to this, the system could be frustrating and not clearly meeting the expectations of the users.

Regarding the acceptance ad-hoc questions results are reported in Tab19.

TAB. 20 ATRAKKDIFF INSTRUMENT RESULTS.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
1. The robotic arm can help to help frail older adults with their daily routines	1	5	2	5	-
2. I perceive the use of the robotic arm can help me to control over the difficulties of the daily users.	4	2	3	4	-
3. This system would make my work easier.	4	3	6	-	-
4. This system would make my work more interesting.	6	2	4	1	-
5. I would use the system because I like to use such appliances.	4	6	-	1	2

6. I would use the system because these appliances are modern.	5	5	1	-	2
7. I would use the system to keep up with the newest technology.	6	4	2	-	1
8. In general, I would like to use this system when it becomes available	4	2	4	2	1

In this case, users opinions are relatively polarized to the neutral/negative pole. 38.46% of the participants stated that the robotic arm could help frail people, but other 38.46% thought that it would be unlikely. If we take into account participants who stated between very unlikely and unlikely to the usability/acceptability questions:

- About half of the participants (53.85%) stated that the system will not make their work easier. The rest of them considered it neutrally (46.15%).
- 61.54% thought that the system would not make the work more interesting.
- Only near a quarter of the sample thought that they would use the system because they like it (23.08%)
- 76.92% would use it because it is modern or because they like to keep up with the new technologies.
- Near to a quarter of the sample (23.08%) would use the system when it would become available, 46.15% would not use it and 30.77% are neutral.

5.3 COMPARISON BETWEEN FIRST AND SECOND PROTOTYPE

We have compared both subjective and objective indicators of the user experience between the first and the second SRS UI-PRO prototype. The subjective experience was assessed by comparing results of the AttrakDiff questionnaire. On the other hand, objective experience was assessed by comparing frequency of failures of the system, time needed to perform the tasks, errors of the user performing the task and hints needed by the user to accomplish the task objectives.

AttrakDiff is a questionnaire which aims to assess the user's feelings about the system. In AttrakDiff questionnaire, both hedonic and pragmatic dimensions of the user experience are studied with semantic differentials. We have clustered ad hoc the items of the AttrakDiff scales based in the semantic content of each item in order to estimate the following constructs: Attractiveness, Insight, Novelty, Stimulation, Reliability and Efficiency. Each of the constructs comprises the following items.

- Attractiveness: annoying/enjoyable, good/bad, unlikeable/likeable, pleasant/unpleasant, attractive/unattractive, friendly/unfriendly.

- Insight: not understandable/understandable, easy to learn/difficult to learn, complicated/easy, clear/confusing.
- Novelty: creative/dull, inventive/conventional, usual/leading edge, conservative/innovative
- Stimulation: valuable/inferior, boring/exciting, not interesting/Interesting, motivating/demotivating
- Reliability: unpredictable/predictable, obstructive/supportive, secure/unsecure, meets expectation/does not meet expectation.
- Efficiency: fast/slow, efficient/inefficient, practical/unpractical, organized/cluttered.

According to these clusters, we can observe the differences in the means obtained between the first and the second trials.

TAB. 21 ATRAKKDIFF CONSTRUCTS RESULTS COMPARISON BETWEEN FIRST AND SECOND ITERATION OF TESTS.

	First Trials		**Second Trials**	
	M	(Sd)	M	(Sd)
Attractiveness	30,6	6,54	23,1	4,51
Insight	19,9	4,70	15,8	4,59
Novelty	23,3	3,33	19,8	3,26
Stimulation	21,2	4,05	3,8	1,17
Reliability	19	3,80	15,4	4,60
Efficiency	18,8	4,49	11,6	3,53

We can observe that the mean score in all the areas have decreased since the first trials. The second system, closer to the real functionalities, is slower than the first one that was probe of concept oriented but probably not based on the real functionalities and robot behavior. The new interface being slower is less accepted by the users than the previous one: Is considered less attractive (25%), less easy to use (21%), severally less stimulating (82%), and less efficient (38%). On the other hand, the lesser novelty experienced could be due to the previous participation in the trials (15%). Taking into account that the sample that took part in both trials is n=10 participants, the results should be interpreted carefully and cannot be taken as significant differences due to the low number of users involved, but as indicators of the usability.

Regarding the objective measurements, the same three tasks of grasping, leaving and moving that were designed in the first trial were assessed in the second trial. In the following tables we can observe differences in the objective measurements: task completion, time, errors, questions asked, hints given, and facial expression.

TAB. 22 GRASPING SCENARIO USER TASK PERFORMANCES, COMPARISON BETWEEN FIRST AND SECOND ITERATION OF TRIALS.

Grasping task	First Trial	Second trial
Task Completion	11 out of 13 completed the task.	13 out of 13 completed the task.
Time	Mean= 92.16 seconds, st. dev. = 33.9 seconds. The fastest of the users completed the task in 34 seconds meanwhile the slowest of the participants that completed the task needed 159 seconds to complete the task.	Mean= 325.85 seconds, st. dev. = 129.57 seconds. The fastest of the users completed the task in 210 seconds meanwhile the slowest of the participants that completed the task needed 600 seconds to complete the task.
Errors	0.53 errors on average were made by the users. Users press the grasping button previously to the robot achieve the correct grasping position.	No errors were made by the users.
Questions asked	0.92 questions on average	0.30 questions on average.
Hints given	1.38 hints were given on average.	No hints were needed to perform this task.
Facial expression	The expression of the participants was neutral in most of the cases no one seemed to be frustrated, scarred or angry about using the system.	The expression of the participants was neutral o positive in most of the cases. Just two participants seemed to be frustrated about using the system.

TAB. 23 LEAVING SCENARIO USER TASK PERFORMANCES, COMPARISON BETWEEN FIRST AND SECOND ITERATION OF TRIALS.

Leaving Task	First Trial	Second trial
Task Completion	All of the participants completed the task.	All of the participants completed the task.
Time	Time: Mean= 125.69 seconds st. dev. = 44.57 seconds. The fastest of the users completed the task in 31 seconds meanwhile the slowest of the participants that completed the task needed 187 seconds to complete the task.	Mean= 302.53 seconds st. dev. = 93.22 seconds. The fastest of the users completed the task in 186 seconds meanwhile the slowest of the participants that completed the task needed 513 seconds to complete the task. This was the time needed by the user and the system to perform the whole task. On the other hand, just M=3.15 seconds were needed by the users to

		perform their part.
Errors	0.38 errors on average.	0.30 errors on average were made by the users.
Questions asked	1.07 questions were asked on average.	0.46 questions on average
Hints given	1 hint was given on average	0.38 hints on average
Facial expression	11 out of 13 were neutral regarding the task, but 2 users expressed annoyance.	The expression of the participants was neutral o positive in most of the cases. Just two participants seemed to be frustrated about using the system.

TAB. 24 LEAVING SCENARIO USER TASK PERFORMANCES, COMPARISON BETWEEN FIRST AND SECOND ITERATION OF TRIALS.

Moving Task	First Trial	Second trial
Task Completion	All of the participants completed the task.	11 out of 13 completed the task.
Time	218.15 seconds, st. dev. = 133.88 seconds. The fastest of the users completed the task in 89 seconds meanwhile the slowest of the participants that completed the task needed 567 seconds to complete the task.	Mean= 202.76 seconds st. dev. = 33.9 seconds. The fastest of the users completed the task in 100 seconds meanwhile the slowest of the participants that completed the task needed 480 seconds to complete the task.
Errors	1.35 errors on average.	Just one user fail in one occasion.
Questions asked	1.38 on average	0.46 questions on average.
Hints given	1.45 on average	1.38 hints were given on average.
Facial expression	The expression was neutral for most part of the users, just one user expressed frustration about controlling the robot.	5 out of 13 users showed frustrated feelings when performing the task

As we can see in the previous tables, objective measurements show the system improvement in the usability related measurements. Regarding task completion all of the users finished the leaving task, while only two failed to finish the grasping task in the first trial and the moving task in the second one.

Regarding time needed to complete the task, the functionalities of the robot such as grasping and leaving have increased the time performance in the final prototype but decreased the time needed to move the robot.

The errors made by the users while performing the task have decreased in all the tasks from.53 to 0 in the first task, 0.38 to.30 in the second task and 1.35 to.07 in the third task.

Less questions were asked in the second trials:.92 to .30 in the grasping scenario, 1.07 to.46 in the leaving scenario and 1.38 to .46 in the moving scenario.

Regarding hints given by the researchers: 1.38 hints were given on average in the first trial, while no hints were needed to perform this task in the second one, 1 hint was given on average in the first trial and 0.38 hints in the second one, and the hints decreased from 1.45 to 1.38 in the second trial.

The results show an improvement in the system usability with the exception of time performance.

5.4 DESIGN SUGGESTIONS BASED ON THE SECOND TRIAL

Based on the feedback obtained in the tasks, few suggestions to be taken into account for future development would be, not only in SRS but also for similar studies, are reported in Tab. 25.

TAB. 25 DESIGN SUGGESTIONS BASED ON VISUALIZATION TRIALS RESULTS

Usability and usability flaws	Design Suggestions
Screen Appearance	
Interface is not intuitive enough.	The interface should only display information concerning to the user. The interface could display tooltips to ease the use.
Gazebo camera is not intuitive enough.	More real vision is encouraged.
	The user should be allowed to place an axis starting point from which start the camera movement
Real life objects are not easy to be distinguished from one another.	If real vision could not be developed. Labels by code of colors and names should be generated.
There is no feedback enough about the task performance.	There should be feedback on the robot task performance. An “uploading” kind bar, percentages or similar indications should be displayed.

The display menu is complicated.	The display menu should be easier, with less steps, avoiding intermediate functional steps.
Rviz and Gazebo make difficult the manipulation due to split attention	Rviz y gazebo should be integrated.
The steps seemed to be too much long and not intuitive.	Minimize the number of steps needed from the user.
Some participants have problems to understand the axis and how does the robot movement works.	The step and the options needed to perform the task should be done automatically. If not higher formation could be needed.
Some participants cannot see the kitchen counter in the Rviz.	The simulation could need more development.
The interface is not all in Spanish.	The interface should be in the language of the country.
The room cannot be seen completely.	Make possible to see the room.
Some part of the interface is in red and blue colors.	Display the options in not complementary colors for the blind impaired.
Movement of the robot	
The camera movement cannot allow completing the task or reaching the object.	-Develop automatic grasping based on labels that could be detected by the robot sensors. -Make the camera mobile making possible for the user to aim to the more difficult situations.
The movement of the robot is too slow, making impossible to react as emergency.	Develop faster response systems. If it is not possible. Give option to an emergency external protocol (ambulance, etc.)
The movement of the robot is too slow, being boring and frustrating for the user.	Develop faster response systems. If it is not possible, minimize the user interaction in order to the user can perform other tasks while the system is working. In this case the user could be notified when

	actions were required.
There is no feedback on the movement of the robot when it hits obstacles.	The system should give feedback about the movement. Some possibilities could be a risk notification (changing % of hit chance, measured by the sensors in real time), notifications when the robot is stopped by an unexpected event.
Having the button pressed to perform the maneuver is tiring and uncomfortable.	The ideal thing should be to click in some point of the map and the robot went there.
The cameras did not see the obstacles making easy to collide.	Integrated camera and map could solve the issue. Automatic obstacle recognition should be improved.
The turning direction in the blue circle changes when 180º are overtaken	Make possible to perform the whole movement. There should be buttons with predetermined turning options (45º, 90º, 180º)
Grasping functionality	
The participant does not know in which part of the grasping sequence is the robot.	It could be a grabbing notification in a “loading bar” way, something that could give feedback about the process is still working and that didn’t get blocked”.
If there is any system error while performing the grabbing the user is not able to know it.	There should be an alarm notification with some instructions of problem solving if it is the case.
The task performance is slow and it is frustrating for some participants.	The task performance should be faster or not require the whole user attention during the task.
The tools are not accessible enough	Select the most important tools and place in a highly visible place (center-up), develop intuitive icons.
Some users prefer to use the circle because is much precise.	Letting the two options for maneuver could be useful. The system should be

There are too many options to perform a single movement.	customizable, allowing the user to choose his interface depending on his abilities.
The display interface is not much intuitive.	Minimize the display to the minimum options. Develop a point and click manipulation could be helpful.
The user did not understand the interactive markers instructions.	Develop more detailed explanations.
The displays window is hard to understand and interact.	Display all information in one window and different tabs for the actions that require it.

5.5 DISCUSSION

The system has been improved since the first presentation to the users, but the new developments and the real time performance of the robot shown in the trials have arisen new circumstances and further improvements needed.

Objective measurements made during tests show the system improvement in the usability related measurements. Regarding task completion all of the users finished the leaving task, while only two failed to finish the grasping task in the first trial and the moving task in the second one; regarding time needed to complete the task, the functionalities of the robot such as grasping and leaving have increased the time performance in the final prototype but decreased the time needed to move the robot; the errors made by the users while performing the task have decreased in all the tasks; less questions were asked in the second trials; regarding hints given by the researchers they decreased too.

The results show an improvement in the system usability with the exception of time performance.

From the users' experience point of view, mainly, the slow speed of robot tasks' performance is frustrating for the users. The changes in the speed of performance of the robot since the first session of tests were due to the technical developments carried out up to the current iteration: the number of functionalities of the robot was grown, the robot can automatically detect the object, gets near to the object in an optimal safe position in order to calculate the trajectory, etc. All of these processes have a computational cost that is reflected in the execution time.

The interface shows a slow navigation of the robot, that is near to the robot real programmed speed of movement. Taking into account that one of the main objectives of SRS is ensure the safety of the users at all times, the speed of performance probably cannot be improved yet. The speed shown in the visualization trials is a safe maneuver speed as observed in the previous trials.

In conclusion, the system interface has been improved since the last trials but the real speed of the robot could be frustrating to the professional users. Further developments should take into account improving safety in the fastest movements, and making possible that the system performs automatically the most part of the tasks to avoid frustration and tiredness in the users.

6 UI-PRO 3D SEMI-AUTONOMOUS REMOTE MANIPULATION AND NAVIGATION TESTS

6.1 INTRODUCTION AND OBJECTIVES

In this evaluation, we tested the part of the professional user interface (UI-PRO) developed by SRS partner BUT as part of the EEU extension. This development work started later than the work on the other user interfaces. For this reason, only click-dummies had been tested before. Our goals were to obtain user opinions and ratings, determine usability issues, and assess the overall utility and suitability of the solution for remote navigation and manipulation by operators in a call center.

The central innovative aspects of this user interface are described in the following paragraphs.

6.1.1 Integrated Sensor Information In 3D Scene:

The user interface employs an integrated view of sensor information and other visualizations. The following figure shows the integrated elements: Laser scanner data (red dots/lines), 2D floor map (grey with black lines, based on laser data), Kinect live RGB and depth data (within yellow “field of view” lines in front of robot), RGB video (lower right), global 3D map (around the robot; explained in next subsection), collision warning indicator (left of robot, lights up whenever robot is close to walls or obstacles), robot footprint indicator (yellow box around robot; this is the collision-relevant safety space).

Users can freely adopt any perspective (e.g. bird’s eye view as in this figure or behind robot, looking at robot from the front, etc.) depending on the requirements of the situation.

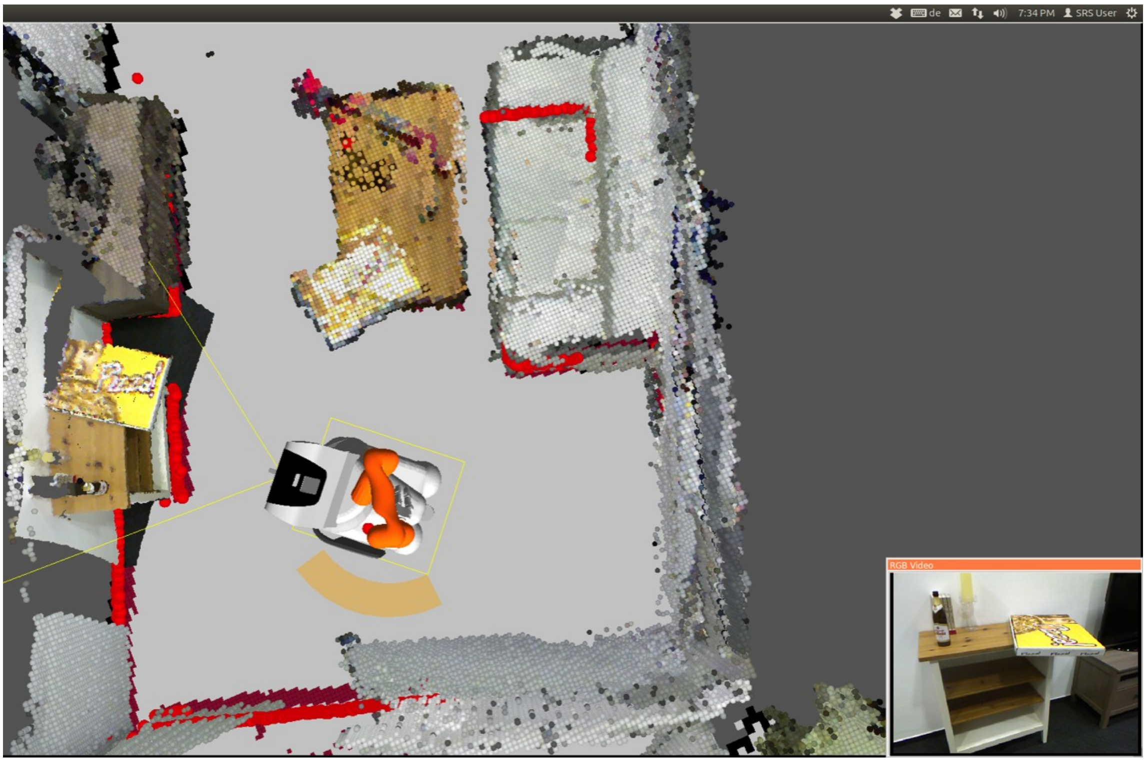


FIG. 40 INTEGRATED SENSOR INFORMATION IN 3D SCENE; FREELY ADJUSTABLE PERSPECTIVE AND ZOOM LEVEL

6.1.2 3D Environment Modeling (Global Voxel Map)

While the robot drives around, it integrates all collected RGB and depth data acquired by the Kinect sensors and combines it into a global 3D map. The user can visualize this map for improved awareness of the environment during navigation and manipulation tasks. The 3D map gets updated during runtime (e.g., objects that were moved get deleted; newly placed objects get added) while the robot drives around.

FIG. 41 THREE-DIMENSIONAL ENVIRONMENT MAP OF AN APARTMENT GENERATED FROM INFORMATION ACQUIRED BY THE ROBOT'S SENSORS WHILE DRIVING AROUND

6.1.3 Semi-Autonomous and Manual Remote Navigation

The user can drive the robot remotely while the robot autonomously avoids collisions (can be turned off for fully manual mode). This operation mode can be used to navigate the robot whenever autonomous navigation fails, e.g. to navigate through narrow passages or when delicate objects are in the proximity of the robot. Users can switch between operation based on the user perspective (current position of virtual camera based on current 3D scene rotation) and based on the robot perspective (similar to conventional joystick-based control).

There are two basic operation modes:

a) In-scene navigation control: The user can either drag a yellow disc and the robot follows it or use arrows to move the robot and a blue wheel to rotate the robot.

b) Space-Navigator: This 3D interaction device combines the arrow- and circle-based movement in one device.

Space-Navigator-based interaction mode was used in this evaluation.

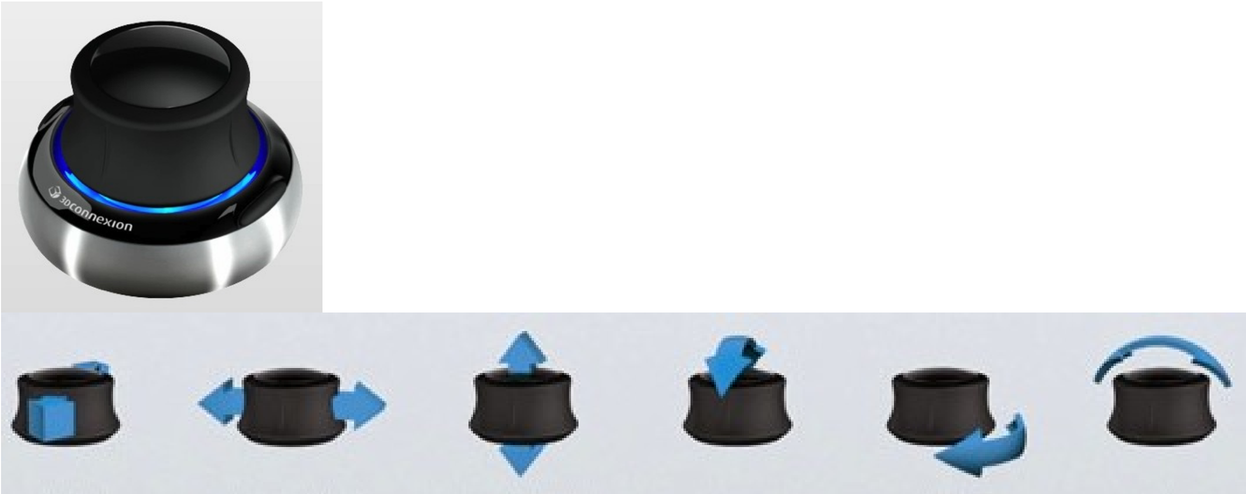


FIG. 42 SPACENAVIGATOR 3D INTERACTION DEVICE FOR REMOTE NAVIGATION AND MANIPULATION

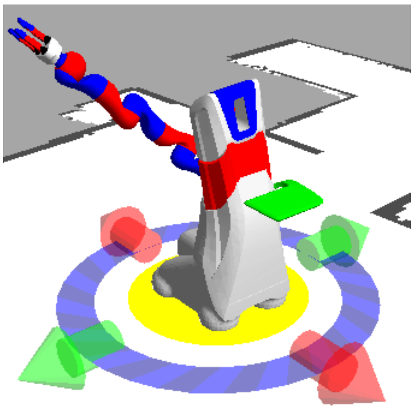


FIG. 43 IN-SCENE NAVIGATION CONTROL

6.1.4 Semi-Autonomous Remote Manipulation

Semi-autonomous remote manipulation is meant to be used in any situation when autonomous manipulation fails to complete or is not feasible from the beginning. Situations can be unknown objects, cluttered scenes, or difficult grasping situations. The User first specifies an area to be removed from collision map by moving a bounding box over the object. The user then specifies the gripper's target position in 3D space. In the last step, the user lets the system simulate the planned trajectory to verify that there are no collisions and the final position is suitable for grasping. After this assistance by the user, the robot then executes the movement. The following figures show this operation mode.

FIG. 44 REMOTE SEMI-AUTONOMOUS MANIPULATION STEP 1: USER SPECIFIES SIZE, POSITION, AND ORIENTATION OF BOUNDING BOX (BOX IS PLACED OVER THE TARGET OBJECT, IN THIS CASE A BOOK IN A SHELF)

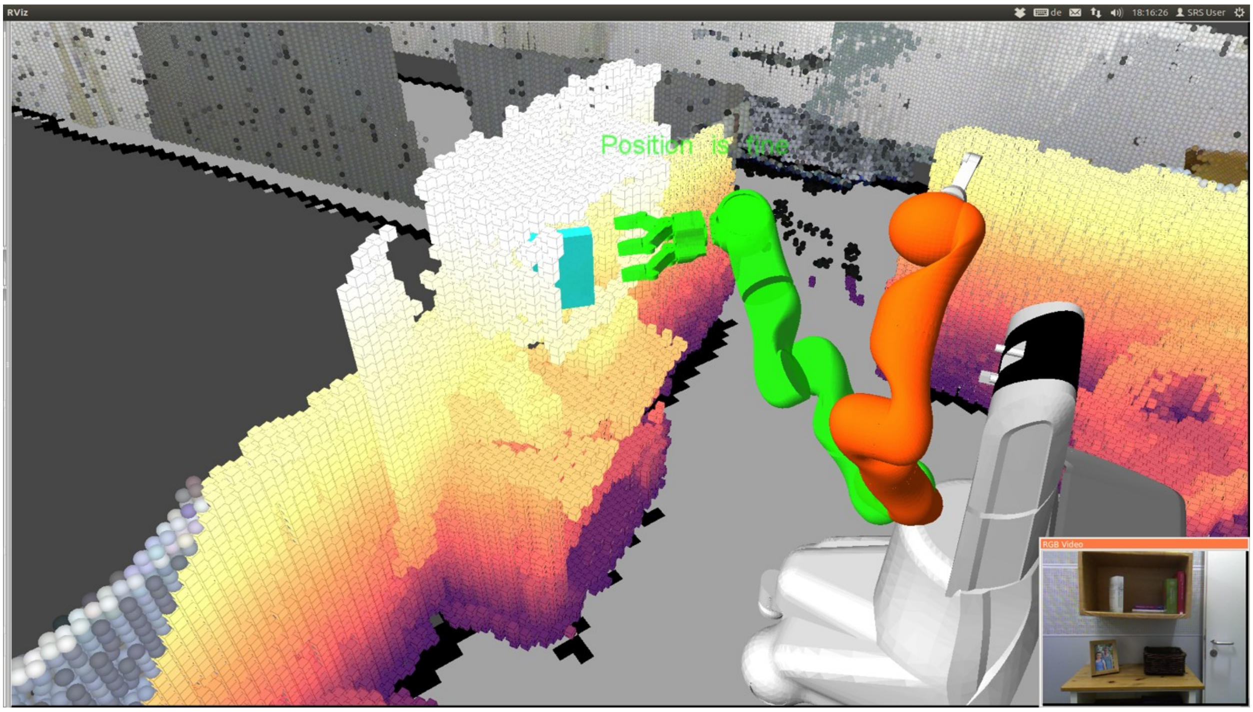


FIG. 45 REMOTE SEMI-AUTONOMOUS MANIPULATION STEP 2: USER SPECIFIES GRIPPER TARGET POSITION

6.1.5 STEREOSCOPIC VISION

As an option, the user interface offers stereoscopic display. This is implemented using NVidia Vision 2 technology and was activated in this evaluation.

FIG. 46 PARTICIPANT AND INTERVIEWER USING STEREOSCOPIC GLASSES

6.2 TESTING SITE

The experiments were carried out at Fraunhofer IPA in Stuttgart, Germany. Office rooms and corridors were cleared of all furniture and around 80 pieces of household furniture and objects were added. Three testing rooms were available plus a separate room on the same floor for the remote operator (participant) and main interviewer, a space for a second interviewer, and a room for technical supervision.

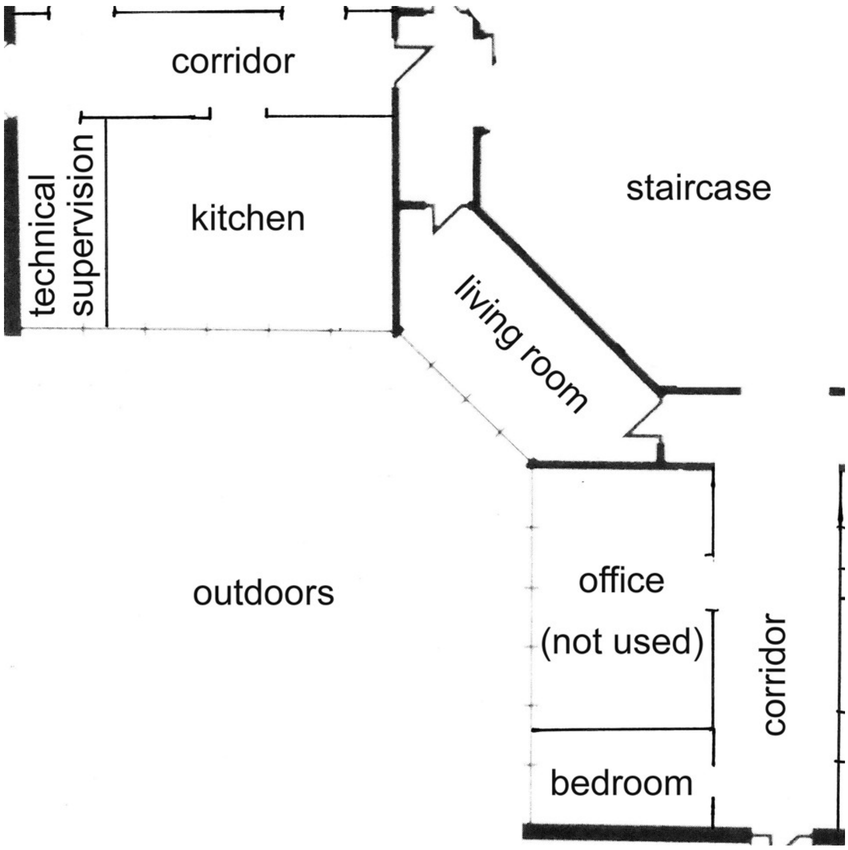


FIG. 47 ROOM PLAN OF TESTING SITE

6.3 PARTICIPANTS

14 people participated in the study. Because the profession of a remote assistant for domestic service robots is not yet existent, participants were recruited based on the assumed professional profile required for future remote operator staff. Participants were aged between 20 and 37 (average age 26.5), had high computer experience (min. 5 years and 10 hours computer usage per week), in most cases completed or ongoing university education, but no robotics-related education or experience.

6.4 PROCEDURE AND USER TASKS

The following table provides an overview of the test procedure.

TAB. 26 OVERVIEW OF THE TEST PROCEDURE

Introduction Phase (secondary interviewer) Welcoming, informed consent, incentive Vision tests Introduction to robot and user interfaces SpaceNavigator training Navigation training in simulation Manipulation training in simulation

Main Phase (participant in remote operator room with main interviewer)
Kitchen apple juice manipulation task:
Bounding box placement
Arm navigation in and grasping

Navigation task:
Living room narrow passage with obstacles

Bedroom book-in-shelf manipulation task:
Bounding box placement
Arm navigation in and grasping

Recapitulating interview

Conclusion Phase (secondary interviewer)
Post-experiment questionnaire

Users accomplished two manipulation tasks and one navigation task. Manipulation tasks included the subtasks of bounding box placement and gripper placement. The first manipulation task (kitchen) was to remotely grasp an apple juice box lying in a non-ideal position in a cluttered scene after the robot dropped an object. The navigation task (living room) was to navigate the robot through a narrow passage with high and low obstacles. The second manipulation task was to grasp a book in of a shelf, which can cause problems for autonomous mode due to one object being inside of another object. The following pictures show the situations at the beginning of each task.

FIG. 48 ROBOT AT START FOR KITCHEN REMOTE MANIPULATION TASK: USERS REMOTELY NAVIGATED THE ARM TO THE APPLE JUICE BOX AND THEN GRASPED IT. A TRAJECTORY OUT WAS THEN PERFORMED TO VERIFY THE GRASP.

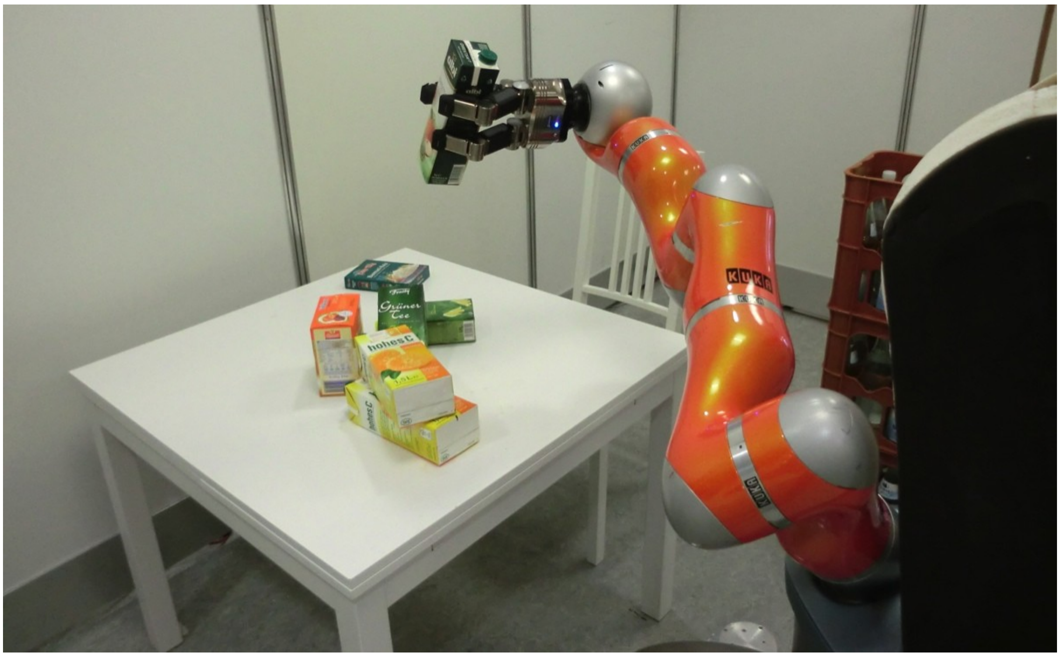


FIG. 49 APPLE JUICE IN GRIPPER AFTER REMOTE GRASP

FIG. 50 ROBOT AT START FOR LIVING ROOM NAVIGATION TASK: USERS REMOTELY NAVIGATED THE ROBOT THROUGH SEVERAL OBSTACLES (PIZZA BOXES, DUST MOP, TRAY, CARTON)

FIG. 51 ROBOT AT START FOR BEDROOM REMOTE MANIPULATION TASK: USERS REMOTELY NAVIGATED THE ARM TO THE LEFT BOOK AND THEN GRASPED IT. A TRAJECTORY OUT WAS THEN PERFORMED TO VERIFY THE GRASP.

6.5 RESULTS

Summative results for AttrakDiff and for a revised version of the questions on system usefulness developed by Don Gnocchi Foundation during previous SRS evaluations are reported in this subsection along with issues detected in the user interface.

Overall, users were able to complete manipulation and navigation tasks in 100% of the cases, meaning that the user interface was highly effective at accomplishing what it was designed for. Analysis of task durations and collisions for the navigation task were still ongoing at the time of writing but results of the pragmatic dimension (PQ) of Mini-AttrakDiff provide an indication of task performance.

6.5.1 AttrakDiff results

We employed the Mini-AttrakDiff instrument, which is a shorter version of the AttrakDiff instrument used in most previous SRS evaluations. Like AttrakDiff, it measures four dimensions of the user experience:

1. Perceived pragmatic quality (PQ) – usability (effectiveness and efficiency)

2. Perceived hedonic quality (HQ)

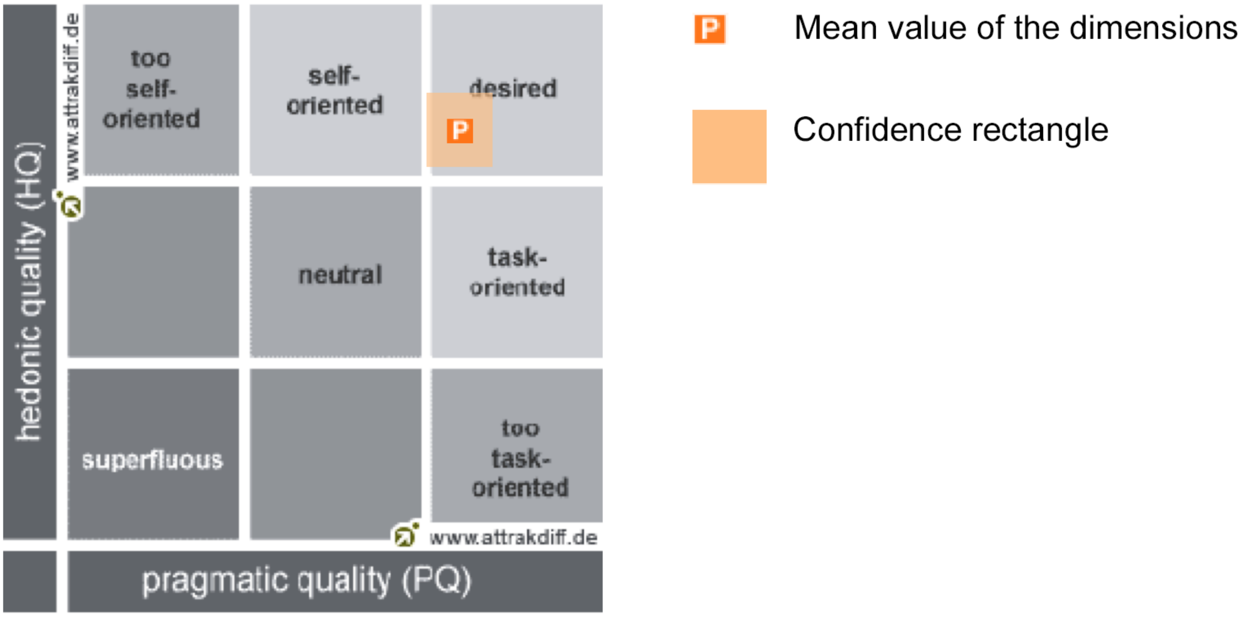
   a) Perceived hedonic quality – stimulation (HQ-S): This quality is based on the effort of people to grow in their knowledge and personality and to become better. The underlying motive is curiosity.

b) Perceived hedonic quality – identity (HQ-I): The product is able to emphasize and complete the users' personality. The product helps the user to express belonging to a specific group of people.

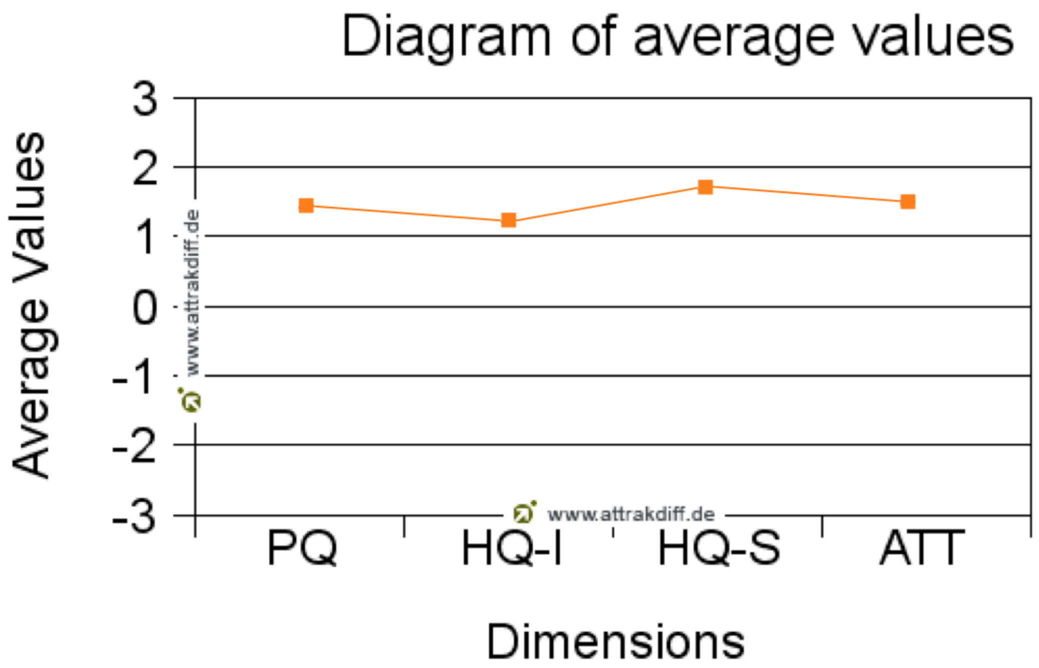
3. Attractiveness (ATT)

Based on the perceived qualities, users form an overall judgment of the product ("good" or "bad"). This overall rating is measured as „attractiveness".

TAB. 27 ATTRAKDIFF PORTFOLIO VISUALIZATION WITH HEDONIC AND PRAGMATIC DIMENSIONS



TAB. 28 ATTRAKDIFF DIAGRAM OF AVERAGE VALUES



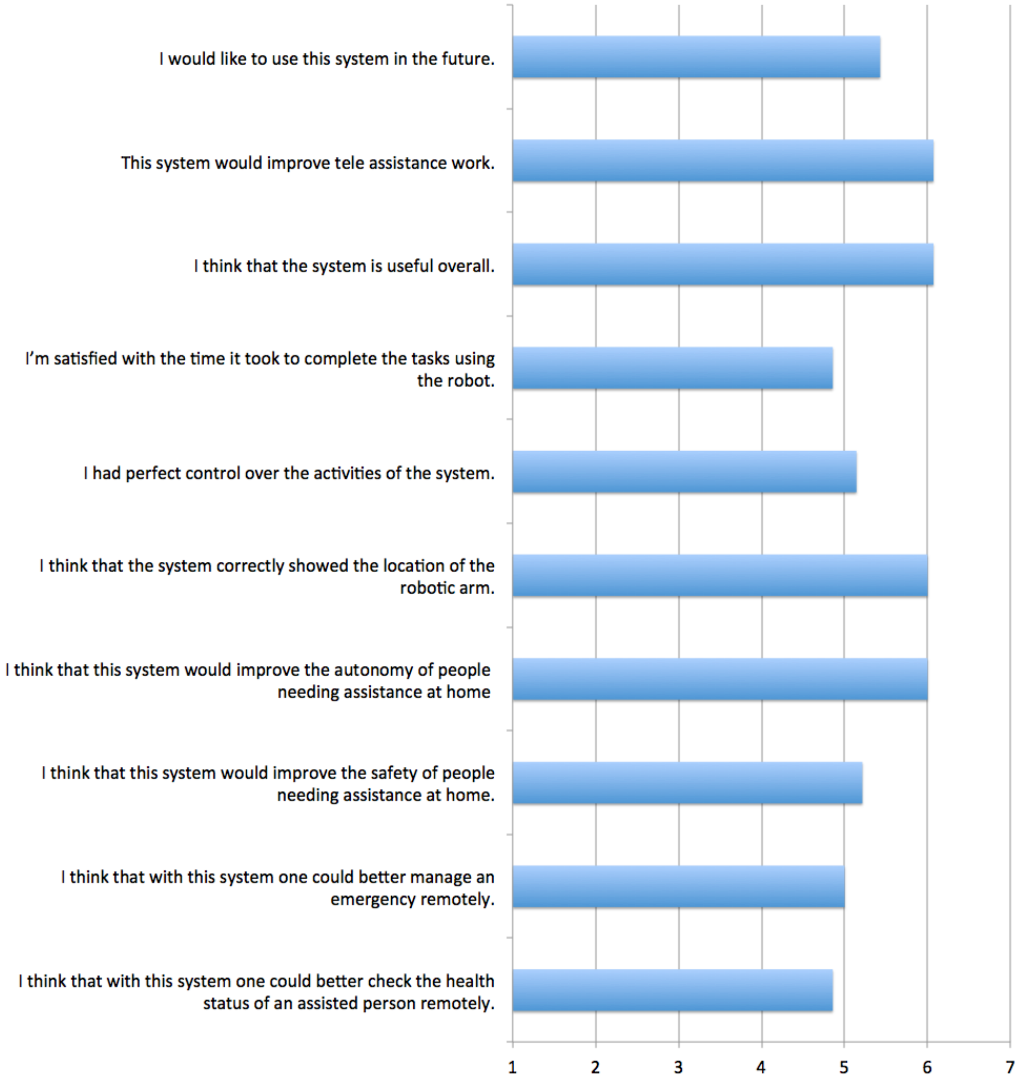
As illustrated in the AttrakDiff portfolio, the user interface overall falls into the “desired” range, which can be considered a good result. The average user rating for usability (pragmatic quality) was 1.4 on a scale of -3 to +3. The average rating for hedonic quality was 1.5. The confidence rectangle is rather small, which indicates good stability of the average values.

The diagram of average values shows that users particularly rated the user interface to be stimulating (HQ-S; M = 1.7) but values for usability (PQ; M = 1.4), identification (HQ-I; M = 1.2), and overall attractiveness (ATT; M = 1.5) are rather high as well. This result indicates that users have a positive overall impression of the user interface. However, with average ratings between 1 and 2, there is still potential for improvement.

6.5.2 Results on usefulness

The following results were obtained for post-study questions on the usefulness of the system (i.e., user interface and robot as a whole).

TAB. 29 RESULTS ON USEFULNESS QUESTIONS; RATINGS 1 “COMPLETELY DISAGREE” TO 7 “COMPLETELY AGREE”



Results on the system's usefulness are encouraging and show a vast improvement over the previous study in Milan where participants used an early version of the UI-PRO user interface with components of partners ROB and BUT (see Section **Error! Reference source not found.**). All average ratings are well above neutral (4) with some reaching close to maximum values. This indicates that users find the system very useful.

6.5.3 Issues Found

Only few usability issues and other user interface issues were found. The area with most issues was positioning and resizing of the bounding box in the manipulation tasks. As the arrows for resizing the box were quite small, some users had trouble hitting them with the mouse, especially using the stereoscopic user interface but also in a condition (not reported in this Deliverable) with monoscopic presentation. Also, these arrows were sometimes covered by voxels of the scene. The rotation wheel of the bounding box in some cases caused similar problems as it was partially covered by the box itself unless the box was resized first. A minority of users seemed to have trouble using Space-Navigator despite the training. We assume that additional training would help resolving this issue. On the navigation task, some users thought that the robot moved too slowly or didn't properly react to their interaction with Space-Navigator. This was due to the thresholds set for collision avoidance and due to the passage to navigate through being very narrow (the robot slows down when close to obstacles). The issues found can be fairly easily remedied in the next version of the user interface.

6.6 DISCUSSION

The result of this evaluation is positive overall. User ratings for pragmatic and hedonic qualities as well as for the system's usefulness are encouraging. Solving the issues identified may further improve AttrakDiff results for usability. Improving hedonic aspects beyond the point they have reached may be challenging, as there are no ready-made recipes for improving users' identification with a system or the stimulation they get from the usage experience. However, we do not see a strong need for further improving the user interface's hedonic aspects, as pragmatic aspects can be considered primarily relevant in the professional field. The task success rate of 100% is a further indication for the appropriateness of this user interface.

7 CONCLUSIONS

The system was perceived interesting and potentially very useful by all user groups during the succession of tests performed. Acceptance is high and this can be highlighted as the better result obtained in this first set of tests with the integrated prototype. Elderly people and private caregivers were more enthusiastic about the novelty represented by SRS system with respect to professional operators; people belonging to this last group are more aware of current state of development of other comparable products and projects, and so they are not so easily impressionable just from novelty represented by SRS. With the additional iteration of the SRS, people belonging to the last group were eventually impressed with the latest revision of the SRS interface based on Space Navigator and Stereo Vision.

Local users (elderly people and young man) and private caregivers were made aware that the presented robot was, at the time of Milan tests, was a prototype, and not fully completed or ready for market. This is also reflected by the low score obtained by indicator "intention to adopt", that didn't reach success threshold of 3.5. Users negative judgment respect the current state of development, were partially due to usability of interfaces, as well as to the technical problems that, at times and to varying degrees, hindered the successful execution of the test scenario and necessitated involvement of the technical personnel to correct the problem.

Subsequent analysis of the tests and the associated problems showed that they could be classified in the following groups, listed according to their effect on the execution of the tests:

- Usability issues of the interfaces – moderate impact,
- Reliability issues – moderate impact,
- Efficiency issue of the overall system – minor impact.

Using the successful indicators specified in D6.1, in Milan tests 11 out 18 targets was met for the elderly people, 12 out 14 targets was met for private caregivers, only 1 out of 14 targets was met for professional operators.

Technical partners in the consortium addressed and solved the identified problems. Two steps of testing iteration were carried out focusing mainly on UI-PRO, which was the interface that during Milan tests obtained lower scores for successful indicators specified in D6.1.

The two last tests' sessions were led in San Sebastian and in Stuttgart focusing on arm manipulation and UI-PRO visualization and remote navigation and manipulation.

During San Sebastian tests, objective measurements - percentage of task completion, number of errors made, number of questions asked by users, hints needed - showed the improvement in the usability related measurements, with the exception of time performance. From the users' experience point of view, the slow speed of robot tasks' performance is frustrating. The changes in the speed of performance of the robot since the first San Sebastian session of tests were largely due to the increasing number of functionalities of the robot, the robot can currently automatically detect the object, gets near to the object in an optimal safe position in order to calculate the trajectory, etc. All of these processes have a computational cost that is reflected in the execution time.

The result of Stuttgart tests was positive and concluded the project successfully. Users were able to complete manipulation and navigation tasks with a successful rate of 100%.It shows that the SRS user interface was highly effective in accomplishing what it was designed for. Results on the system's usefulness are encouraging and show a vast improvement over the previous study in Milan. All average ratings are well above neutral (4) with some reaching close to maximum values. This indicates that users found the system very useful. User ratings for pragmatic and hedonic qualities are encouraging; by solving the issues identified in the test may further improve AttrakDiff results on usability. Improving hedonic aspects beyond the point they have reached may be challenging, as there are no ready-made recipes for improving users' identification with a system or the stimulation they get from the usage experience. However, based on user feedback, the need for further improving the user interface's hedonic aspects is low, as pragmatic aspects can be considered primarily relevant in the professional field. The task success rate of 100% during the test is a further indication for the appropriateness of this user interface.

8 REFERENCES

Wessa, P. (2012), Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, URL http://www.wessa.net/