ROBOTS IN ASSISTED
LIVING ENVIRONMENTS

UNOBTRUSIVE, EFFICIENT, RELIABLE AND MODULAR
SOLUTIONS FOR INDEPENDENT AGEING

Research Innovation Action

Project Number: 643892 Start Date of Project: 01/04/2015 Duration: 36 months

DELIVERABLE 6.9

User evaluation report

Dissemination Level	**Public**
Due Date of Deliverable	Project Month M9, December 2015
Actual Submission Date	24 December 2015
Work Package	WP6, *Piloting and evaluation*
Task	Task 6.4, *User evaluation*
Lead Beneficiary	FSL
Contributing beneficiaries	NCSR-D
Type	R
Status	Submitted
Version	Final

Project funded by the European Union's Horizon 2020 Research and Innovation Actions

Abstract

This deliverable reports the findings of User Evaluation of the Formative Phase of Pilot Trials. The ultimate goal is to evaluate the *Graphical User Interfaces (GUIs)* used during the pilot in terms of usability of the system. The report includes a description of the measured variables, the analysis methods used, the results, and a discussion section describing the main findings and their implications

History and Contributors

Ver	Date	Description	Contributors
00	05/10/2015	Document structure	NCSR-D
01	27/11/2015	First draft	FSL
02	3/12/2015	Sections 4.2 and 4.3	NCSR-D
03	17/12/2015	Pre-final version	FSL
04	23/12/2015	Internal peer review	NCSR-D
05	23/12/2015	Addresses peer review comments	FSL
Fin	24/12/2015	Final preparation and submission	NCSR-D

Executive Summary

This deliverable reports the findings of User Evaluation of the Formative Phase of Pilot Trials. The ultimate goal is to evaluate the *Graphical User Interfaces (GUIs)* used during the pilot in terms of usability of the system and improvement in quality of life. The report includes, other than results of the statistical analysis of the Formative Phase, an accurate discussion section describing the qualitative evaluation of the interface; it encompass the comments of the elderly users recorded during the experimental protocol and the suggestions of the researchers that conducted the pilot. Finally, at the end of the documents are listed the main findings and their implications to GUI design.

Abbreviations and Acronyms

ADL	Activities of Daily Living
ASQ	After-Scenario Questionnaire
IADL	Instrumental Activities of Daily Living
interRAI	International collaborative to improve the quality of life of vulnerable persons through a seamless comprehensive assessment system. Cf. http://www.interrai.org
interRAI HC	The interRAI Home Care Assessment System
interRAI LTCF	The interRAI Long-Term Care Facilities Assessment System
MMSE	Mini Mental State Examination
PIADS	Psychosocial Impact of Assistive Devices Scale
SUS	System Usability Scale
GUI	Graphical User Interfaces

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Purpose and Scope

RADIO presents a domestic assistant and home automation profile to the end-user, which most importantly acts as an unobtrusive health monitoring system.

RADIO's main objective is an unobtrusive monitoring system whose equipment is an obvious and accepted part of the user's daily life, by adopting a smart home/assistant robot approach, where the sensing equipment actively strives to be obvious and closely located to the user; that is, we propose that robot companions and assistants are used to collect the data needed for medical monitoring.

RADIO system will provide a pool of ICT based in-home services that will be offered to elderly users that live at home to improve time to be spent autonomously at home. Although the RADIO system is primarily presenting a domestic assistant and home automation profile, it is also acting as an unobtrusive health monitoring system and as an instrument for medical evaluation. It will ensure the timely availability of the patients' clinical and behavioral data to allow timely prognosis and clinical actions. Through its direct involvement in end-users daily activities, RADIO observes *activities of daily life* and *mood*. These observations are used to establish patterns and identify deviation. Moreover, RADIO empowers new care service provisioning models based on the remote supervision of the elderly/patients from the medical experts and/or health professionals or care-givers. It deals with the extraction/derivation of reinforced medical knowledge associated with symptoms, good practices, treatments and personalized patterns of treatment for elderly users.

Objectives of the study:

- Measure validity of the Radio system
- Evaluate functional activities and mood
- Improving Quality of Life
- Measure Usability

The work reported here evaluates the usability tests of the *Graphical User Interfaces (GUIs)* user interfaces of S&C's Home Automation systems in order to inform the development of user interfaces for RADIO target group. This report focuses on evaluating the results obtained during the Formative Phase of Pilot Trials that is dedicated to the collection and analysis of the usability of the first version of the RADIO GUIs. More specifically, this report summarizes the methods of the usability tests. It then goes on presenting the descriptive statistics of the data collected and the statistical analysis performed. Moreover, this repost presents the qualitative evaluation of the usability testing. The qualitative evaluation drives the requirements for the design of RADIO's GUIs, which are presented in Section 4.2. Suggestions for improvement for the next pilot phase are discussed at the end of this report.

1.2 Approach

This is a multicenter and multinational non-experimental clinical study. The target population is elderly people who need assistance in order to maintain their independence and quality of life.

The study will be distributed in three phases:

1. Formative phase; first pilot at FSL
2. Intermediate phase; second pilot of RADIO components at FSL
3. Summative phase; final RADIO pilots

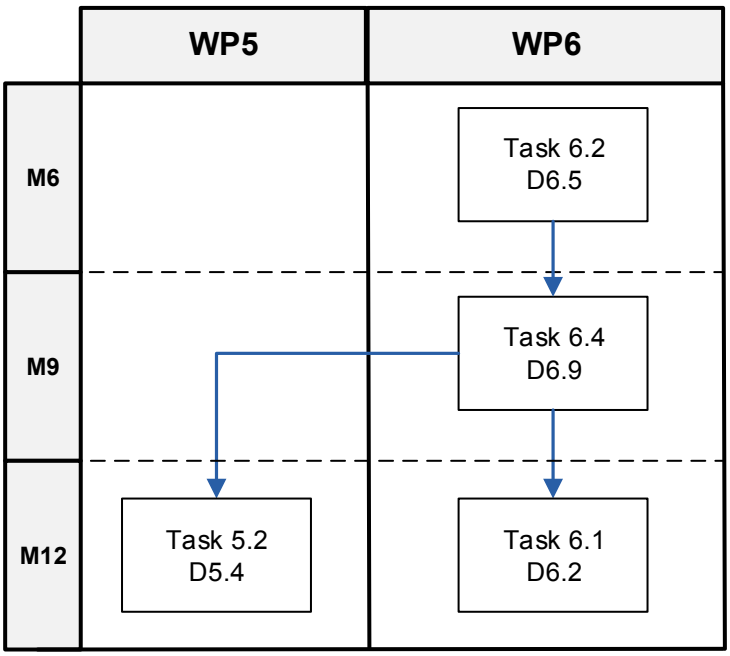

Figure 1. Dependencies between this deliverable and other deliverables.

Formative phase: The first pilot was carried out at FSL premises with elderly end-users. The objectives of this pilot are (a) to provide data for a purely formative evaluation of the usability of existing user interfaces for controlling home automation and (b) to refine the piloting plan itself into its second version.

Intermediate phase: The second round of pilot, also at FSL premises, will be realized with the first versions of user interfaces, devices, and the robotic platform delivered on M12 and M15. The objectives of this pilot are (a) to provide data for the formative evaluation of early RADIO components for usability and fitness for purpose; and (b) to refine the piloting plan itself into its third version.

Summative phase: This final phase includes two sets of pilots, one at FHAG premises and one at the private homes of FZ clients who have volunteered to participate, implementing the third version of the piloting plan. The objectives of these pilots are (a) to validate the prototype of the overall RADIO ecosystem; and (b) to provide data for the final, summative user evaluation report and medical evaluation report.

1.3 Relation to other Work Packages and Deliverables

This report describes the User Evaluation of the Formative Phase Controlled Pilot Trials. These trials were executed at FSL premises during July – September 2015.

This report includes a description of the measured variables, the analysis methods used, the results, and a discussion describing the main findings and their implications for designing the User Interface for the primary user (D5.4). It also suggests refinements for the next piloting plan (D6.2). This report builds upon D6.5: *Controlled Pilot Trials*, which is summarized in Sections 2 and 3. New material in this report is the analysis in Section 4. These dependencies are also shown in Figure 1.

2 METHODS

This section provides a brief description of the participants and the set-up. It goes on with reporting the testing scenarios. It then briefly presents the comprehension and usability evaluation data collected as well as the feedback received from participants during personal interviewing. It goes on with a detailed description of the statistical methods used.

2.1 Participants

The piloting plan (D6.1) foresees a user group of 30 elderly participants satisfying the following criteria:

Inclusion criteria:

- Older than 64 years old
- Ability to walk without human assistance indoors
- Need supervision in almost two IADLs
- Willing to participate in the study and wanting to co-operate in all its parts, accepting the performance regulations and procedures provided by the researchers

Exclusion criteria:

- Moderate/severe mental disease, such as dementia, according to clinical criteria -DSM-IV-TR and MMSE score ≤18 or neuropsychiatric disorders
- Moderate/Severe disability ADL< 4
- Acute medical conditions
- Unable to fully understand the potential risks and benefits of the study and give informed consent. Subjects who are unable or unwilling to cooperate with study procedures.
- Blind, deaf, languages problems

Subject population fulfills the inclusion and exclusion criteria descripted above. Indeed, all subjects are older than 64 years, with a mean age of 72,27 yrs.; cognitive capabilities, evaluated as MMSE, have a minimum score of 19.2 with a mean of 26.2 (see Table 1). Population distribution related to age, sex, education and cognitive capabilities is shown in Table 1.
Regarding the need of supervision in IADL, according to the inclusion criteria (D2.1), the distribution of population for each item of the IADL scale is shown in Figure 2.

Table 1. Demographic data of participants.

	Age	**Education**	**MMSE corr**	**Sex**
Mean	72,27	12	26,2	N/A
SD	5,2	3,7	2,37	N/A
				16F/14M

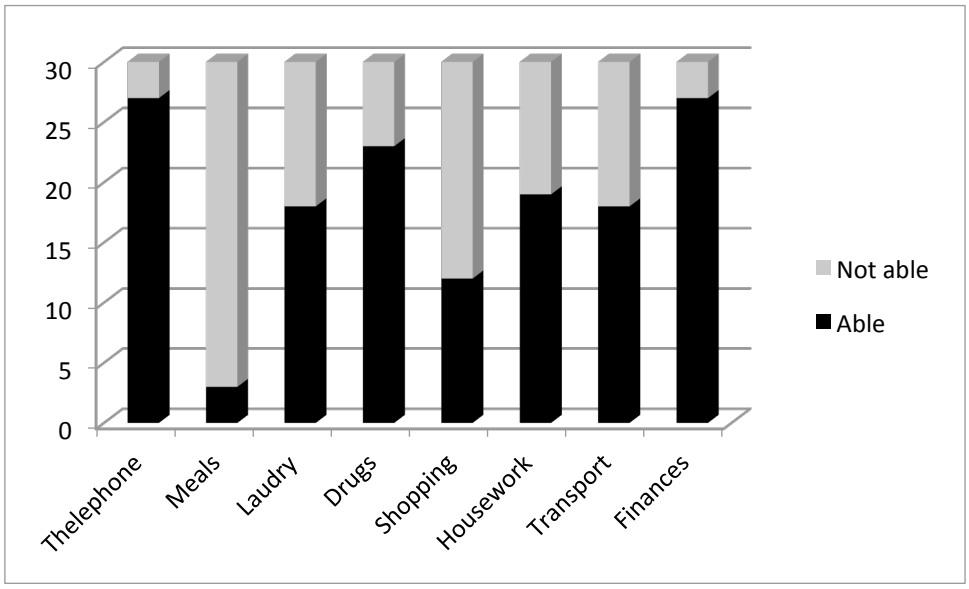

Figure 2. Need of subject population for supervision in IADLs.

The early loss IADLs in our participants were shopping and meal preparation, while most of the users were able to use telephone and manage finances by themselves.

2.2 Set-Up

As described in details in D6.5 the experiment was conducted at FSL premises. In the preparation phase it was equipped with the S&C Smart Home system that allows managing daily home activities. The system included the following products:

- 1 Gateway for communication
- 1 Door/Window sensor
- 1 Multi-sensor 4X1 (Motion, Temperature, Humidity and luminosity)
- 2 Smart Energy Plugs (Electricity measuring and on/off)

Moreover, it encompassed a Tablet, Galaxy Tab A Display 9.7 with EnControl solutions for smart home.

Before starting any experimental test, each user was assessed through multidimensional evaluation by InterRAI (see D2.1), MMSE and IADL. After the baseline assessment the scheduled program consisted of 2 sessions within a week for each participant lasting an average of 1 hour.

During the day before the experimental session researchers showed to the subject the smart home environment and tablet with S&C service. The subject was encouraged to familiarize himself/herself with the test interface and go through this interface at least 3 times.

In the first session FSL team asked him/her to perform some actions in order to access home automation functionality trough the tablet, asking them to perform a pre-defined protocol as already defined in D6.1 and D6.5.

The protocol included the following actions:

1. Arm the system (intrusion detection)

2. Switch on a lamp

3. Switch off a heater

4. Verify if a burglar is moving in the home

5. Schedule the lighting of the lamp and heater and locking at a predefined time

6. Verify electricity consumption

7. Verify alarm message

During the scenario, the researchers watched carefully and took notes about how users interact with the website to see if the interface is inhibiting users from accomplishing their desired goals.

In order to verify the acquisition of the first session user came back after two days. During this phase researchers asked him to try again the scenario proposed in the previous meeting.

The objective of the second session of the experiments was to evaluate, after a period of non-use, how much a person remembers about the interface and the browsing process.

At the end of the session each users was asked to answer to usability questionnaires, System Usability Scale (SUS), Psychosocial Impact of Assistive Devices Scale (PIADS) and After-Scenario Questionnaire (ASQ).

2.3 Evaluation Variables

At the baseline users were asked for relevant information regarding demographic data (age, sex, birth date, education) and were evaluated on functional status (section G of interRAI LTCF and IADL), mood-behaviour and cognitive aspects (interRAI LTCF specific sections and MMSE).

After the deployment of the experimental sessions each user was assessed regarding usability through System Usability Scale (SUS), Psychosocial Impact of Assistive Devices Scale (PIADS) and After-Scenario Questionnaire (ASQ).

2.3.1 Comprehensive Evaluation

As described in details in D2.2, we use some specific sections of the *interRAI Long-Term Care Facilities Assessment System (interRAI LTCF)* as tool to provide a comprehensive and standardized evaluation of the needs, strengths, and preferences of persons receiving short-term post-acute care in skilled nursing facilities as well as persons living in chronic care and nursing home institutional settings. In the Formative Phase subjects were evaluated, thorough the specific sections of interRAI, ADL, IADL, mood-behaviour and cognitive aspects.

2.3.2 Usability Evaluation

This section lists the instruments that were used to measure the usability variables that are the focus of this study.

2.3.2.1 System Usability Scale (SUS)

The *System Usability Scale (SUS)* is a simple, ten-item scale giving a global view of subjective assessments of usability. It consists of a 10 items questionnaire with five response options for respondents: from strongly agree to strongly disagree. Originally created by John Brooke in 1986, it allows evaluating a wide variety of products and services, including hardware, software, mobile devices, websites and applications.

The System Usability Scale actually covers a variety of aspects of system usability, such as the need for support, training, and complexity, and thus has a high level of face validity for measuring the usability of a system. SUS scores have a range of 0 to 100.

2.3.2.2 Psychological Impact of Assistive Device Scale (PIADS)

It is a scale designed to measure the impact of assistive product of the quality of the subject's life. This questionnaire is composed by self-administered 26-item and it investigate three psycho-social aspects:

- Competence: Measures feelings of competence and usefulness.
- Adaptability: Signifies a willingness to try new things.
- Self esteem: Indicates feelings of emotional wellbeing and happiness

The competence subscale is composed of 12 items related to perceived functional capability, independence, and performance (examples: adequacy, efficiency, and skillfulness). The adaptability subscale is composed of 6 items that reflect inclination or motivation to participate socially and take risks (examples: ability to participate, willingness to take chances, and ability to take advantage of opportunities). The self-esteem subscale is composed of 8 items reflecting self-confidence, self-esteem, and emotional wellbeing (examples: sense of control, happiness, and self-confidence).

PIADS can be used to assess the impact of any assistive device (AD), prosthesis or medical procedure, and can be used with people of all ages and abilities

Each item in scored through a 7-points scale. Scores can range from –3 (max. negative impact) to +3 (max. positive impact).

2.3.2.3 AFTER SCENARIO QUESTIONNAIRE

The *After Scenario Questionnaire (ASQ)* is to be given to a study subject after he/she has completed a normal condition scenario. The user is to circle their answers using the provided 7 point scale (the lower the selected score, the higher the subject's usability satisfaction with their system).

2.4 Statistical Analysis

This paragraph describes the statistical analysis that was used in order to analyze the data.

Data were analyzed with some descriptive statistics, in order to describe the main feature of measures collected and summarize the sample, and inferential statistics, in order to learn about elderly population. We used as descriptive statistics the mean and median as a measure of central tendency and standard deviation as a measure of dispersion of data. We used as inferential statistics Pearson correlation in order to evaluate the positive or negative linear dependency between the two demographic variables (age and the level of global cognition) and measures of impact of technology.

3 RESULTS

This section presents the results of the statistical analysis as already defined in D6.5.

The statistical analysis at this stage of the project has been performed on usability in order to test the GUI, according to the main outcome of this phase of the study.

Usability analysis was performed by a descriptive study of the results of the various questionnaires used for this purpose.

Table 2 shows the descriptive statistics of the usability tests. ASQ has a 7 point scale from 1 to 7 with the score 1 as the most positive score. In the table ASQ has a mean value of about 2 for each of the three questions showing a positive effect.

Regarding SUS, a score above 68 (the average score among 500 studies) is considered as a general positive perceived usability of the tool. The table shows a mean value of 71.5.

Finally, about PIADS, each item is scored through a 7-points scale; scores can range from –3 (max. negative impact) to +3 (max. positive impact). The table shows for each of the three psycho-social aspects a positive impact.

Table 2. Descriptive Statistics of Usability Tests.

	ASQ 1	**ASQ 2**	**ASQ 3**	**SUS**	**PIADS Ability**	**PIADS Adaptability**	**PIADS Self esteem**
Mean	2.47	2.17	2.13	71.5	1.2	1.39	.75
SD	1.41	1.37	1.36	12.36	.44	.83	.48
Median	2	2	2	75	1.16	1.16	.75

Table 3 shows a significant correlation between age and PIADS (ability and self esteem). These results suggest that as the age of the subject increases, the user meets many difficulties to complete the task using radio interface.

Moreover, as shown in table 3, the MMSE does not correlate with usability test in the range score adopted for the study, confirming our choice to include in the study only patients with a MMSE ≥ 18.

Table 3. Pearson Correlation between Age, MMSE and Usability scales.

		ASQ 1	**ASQ 2**	**ASQ 3**	**SUS**	**PIADS Ability**	**PIADS Adaptability**	**PIADS Self esteem**
Age	**Pearson**	.279	.290	.244	-.135	**-.420***	-.125	**-.435***
	Sig. (2-tails)	.135	.121	.194	.476	**.021**	.511	**.016**
MMSE	**Pearson**	-.035	-.198	-.123	.140	.069	.200	-.125
	Sig. (2-tails)	.856	.294	.517	.462	.716	.289	.511

In conclusion results of usability evaluation through the selected questionnaires in the Formative Phase, demonstrated good acceptance and satisfaction rates amongst participants. We can state that we obtained a positive message about the benefits of RADIO user's interface for helping elderly people in real life situations. The technology evaluated appeared simple, reliable and effective and possibly tailored to individual needs.

4 DISCUSSION

4.1 Qualitative Usability Evaluation

While the statistical analysis with usability questionnaires provides an overall evaluation of user's satisfaction and GUI's usability, in this paragraph we report qualitative evaluation collected from the researchers during the experimental sessions. The qualitative evaluation of the Formative Phase consists of the user suggestions recorded during the pilot and researchers examinations during the execution of the scenario. This information was collected in order to add further insights on the usability of the GUI for the users of RADIO system, as lack of confidence in the system may restrict the elders in making use of RADIO system. It allows to discuss central usability aspects such as the readability of the elements on the screen, complexity of the operation system, and to get general feedback about the system and its features.

In the following, we present the difficulties highlighted from the users or observed from the FSL researchers during the formative phase; including screenshots of the relevant GUI panels.

4.1.1 Colors and text

Some users had difficulties to read the information on the display. They said to us *"I cannot see well the buttons"* or *"I cannot read the words written in a small size at the top of the screen"*. These visual problems are frequently observed in the elderly as they have less sensitivity to color contrast than the younger people.

Notably, FSL researchers outlined as there is not sufficient contrast between the icons and the background, both in the main page than in the subsequent screens. As showed in figure 3, brown icons are on a brown background. As pointed out elderly need a clear contrast between the icons and the background and also between the text and the background.

Another example in the Figure 4 where is showed the text with low contrast and a small font-size.

All the users quickly understood the meaning of the icons apart from the "control" icon that is confused with an on/off button. It could be a good option substitutes this icon for example with the plug icon of the specific section for a more intuitive illustration.

4.1.2 Reduction of Complexity

Most of the users showed difficulties to understand the meaning of Security and Control icons/text on the main page (see figure 5). They often asked us to explain again the meaning of the text but they don't understand completely. When in the second session we asked them to explain these two icons they aren't able to do it in most of case. Effectively these icons show other than the specific functionality also other information, such as an active state of the security or scheduling on/off.

Single task per page, not only reduce complexity of application, but also reduce attention load for older users. So, clearly separated task is the factor that may increase usage performance. It could be reasonable to simplify this aspect (one key-one function).

Moreover, items per working page should be limited in order to avoid an overload of information. Including in same panel visual elements that provide sensor readings as well as elements that provide access to functionalities requires more attention and can generate confusion.

For example the main page (see figure 6) shows comfort information (temperature, humidity) other than the four buttons related to system functionalities. So, it could be useful to remove the comfort information from the main page and only provide it in its specific section.

Figure 3. Color contrast

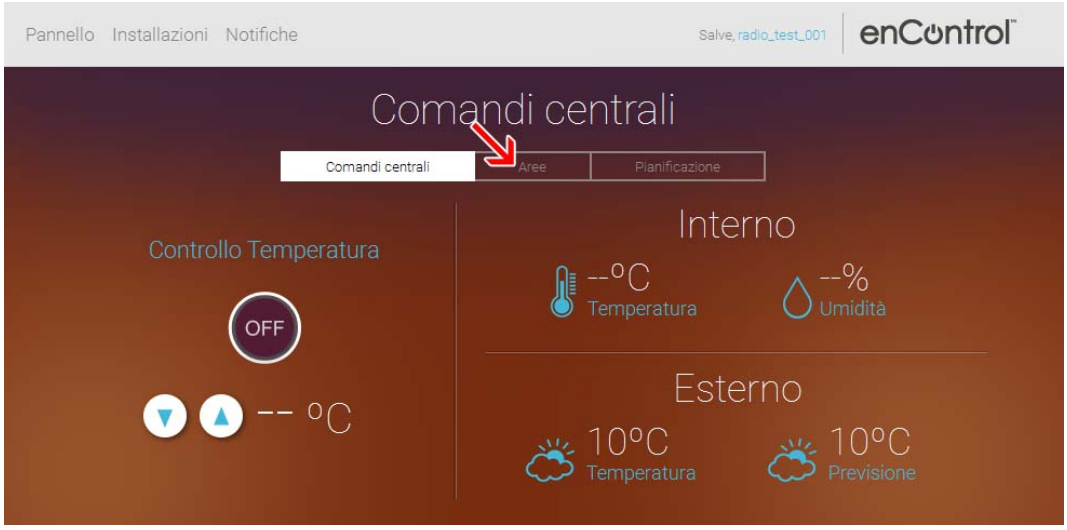

Figure 4. Font size

Figure 5. Icons

Figure 6. Complexity of the main page

Figure 7. Dropdown menu

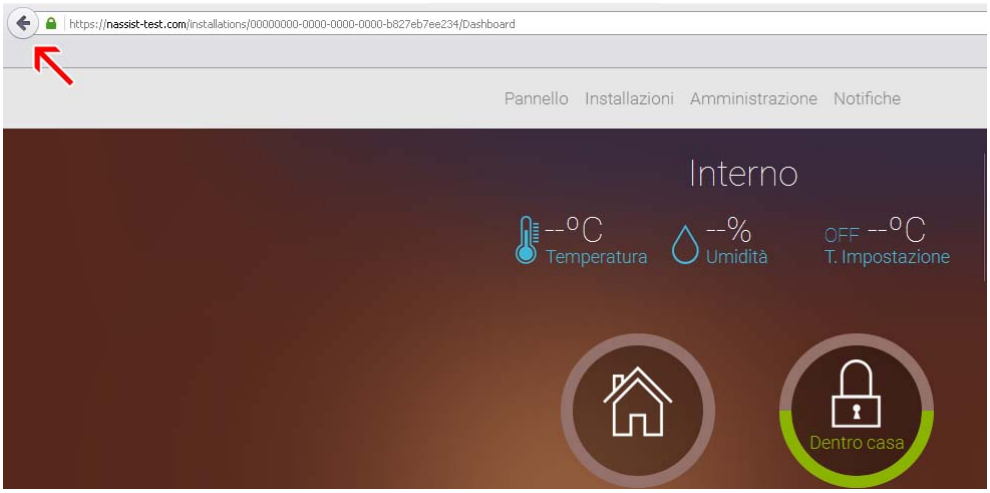

Figure 8. Functionalities

For the same reason, FSL researchers outlined as functions that are rarely used or not necessary should be removed in order to provide simplicity of application. If feasible it should be possible to provide customized facility for elderly users such as adjustable text size or adjustable button size or remove functions if not used.

Also for the dropdown menu (see Figure 7) in the scheduling plan the users stated that *it is difficult to read*. More over the user has to open the drop down menu and select the option. If possible the options should be listed (text or icons) so the user can select directly the required one.

4.1.3 Clear structure of task

For older people, too much multimedia usages create confusions to the utilization of the application itself. Required multitasking operations should not be applied as well. FSL researchers stated as navigation bar on the top of screen contains too much information and too many subtasks (also in this case the contents is difficult to read). For example, in order to came back to the Homepage the user has to click on panel and select an option, while a direct function to return to the homepage could minimize the complexity of the task. At the same time some function such as installation could be removed.

During the pilot, most of the users doesn't find how to return to the previous page; in fact for example in the "scheduling plan", in order to select a page, the user has to go to the navigation bar and select the right option or to use the arrow of the browser at the top of the screen (see figure 8). It could be useful provide an icon in the display in order to simplify this task.

Some users asked us to add, to the camera outside the door, a door phone to talk with people and, if feasible, to open the door.

Moreover, it could be useful to provide a short video with the main features of the system that elderly can see to improve his confidence and solve emerging usability problems.

4.2 Graphical User Interface Requirements

Based on the qualitative usability evaluation, Table 4 collects the key elements that need to be taken into consideration in order to design a friendlier GUI for our target group. These will be used as requirements for the RADIO User Interfaces (WP5).

4.3 Other requirements

This paragraph sets other requirements (not strictly speaking related to usability) that became apparent through these tests.

Using door camera and phone and opening the door introduces technical requirements with respect to the automation hardware that is installed. This requirement will be discussed in the context of the system architecture (Task 4.1).

Table 4: Graphical User Interface requirements based on the qualitative usability evaluation.

Qualitative Attribute	GUI Requirement
Colors and Text	
Color contrast	Intensify color contrasts between the background and icons/text
Size of text	Larger text size
Reduction of Complexity	
Minimize information	Remove items per working page (items not relevant to the task performed at the time).
	Remove functions that are not necessary (e.g. installation item).
Icon Clarity	Change icon to a more intuitive Illustration. (e.g. "Control" icon confused with an on/off button)
Easy access of information	Drop-down lists to be replaced by all-context menus for direct access to the required item.
	Single task per page/ one key-one function.
	Easy access to home page.
	Easy return to previous page. In general, access to the different panels should not use the browser's navigation buttons: all navigation should be foreseen in the GUI design and implemented as GUI functionality.
Clear structure of task	
Simple chain of actions to carry out an operation.	The GUI should guide the user through the steps needed in order to achieve a task. No information or options that is not directly related to the current task should be presented.
	The user should only be given the absolutely necessary options required for this task, and reasonable defaults should be applied in as many cases as possible.
Online help	
Provide a short video with the main features of the system that elderly can see to improve his confidence and solve emerging usability problems.	The GUI should be able to play such short videos, ideally offering videos relevant to the current task.

4.4 Conclusion regarding the piloting plan

The Formative Phase of the study protocol showed that elderly users have a good acceptability and satisfaction of the GUI of the RADIO system in terms of overall usability, as resulting from the statistical analysis of the usability questionnaire included in the study design. Moreover, these results show that the level of cognitive abilities we identified and adopted in the selection criteria as a range score of the Mini Mental State Examination was correct. In fact the MMSE ≥ 18 (MMSE < 18 represent an exclusion criteria) does not correlate with usability, confirming our choice to exclude elderly users with moderate-severe cognitive problems.

Conversely, usability correlates with age showing more difficulties with increasing age. This data is consistent with the findings of the qualitative usability evaluation and with the age-related changes of the aging process.

Observations of elderly users included in the Formative Phase and problems pointed out from the researchers conducting the experiment provide several important feedbacks that have to be taken in consideration for the implementation of the GUI usability. Usability will become even more important in the following phases, when elderly users will use RADIO system every day and the interface will encompass several functionalities. As emerged from this phase it will be important understand which types of problems the elderly users will face off; so we can suggest, also in the following phase, to add a qualitative usability evaluation other than the dedicated questionnaires for usability evaluation.

REFERENCES

[Law69] Lawton M.P. and Brody E.M. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 9, 179-186, 1969

[Folstein75] Folstein MF, Folstein SE, McHugh PR. «Mini-mental state». A practical method for grading the cognitive state of participants for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198

[Jutai04] Jutai J, Day H. Psychosocial Impact of Assistive Devices Scale (PIADS). Technology and Disability 2004; 14:107-111

[Lewis91] Lewis JR. An after-scenario questionnaire for usability studies: Psychometric Evaluation over three trials. SIGCHI Bull., Vol. 23, No. 4. (1991)

[Brooke96] Brooke J (1996) SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL (eds) Usability evaluation in industry. Taylor & Francis, London, pp 189–194