ICT-2009-248730

Florence

Multi Purpose Mobile Robot for Ambient Assisted Living

STREP
Contract Nr: 248730

Deliverable: D6.4 Report on the Testing and Evaluation Methodology for the Living Lab Testing

Due date of deliverable: 31-08-2012
Actual submission date:

Start date of Project: 01 February 2010 Duration: 36 months

Responsible WP: 6 (Novay)

Revision:

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Service	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (excluding the Commission Services)	

0 DOCUMENT INFO

0.1 Author

Author	Company	E-mail
Mark Melenhorst	Novay	Mark.melenhorst@novay.nl
Melvin Isken	OFFIS	melvin.isken@offis.de
Dietwig Lowet	Philips	dietwig.lowet@philips.com
Chrissy van de Wal	Novay	Chrissy.vanderwal@novay.nl
Henk Eertink	Novay	Henk.Eertink@novay.nl

0.2 Documents history

Document version #	Date	Change
V0.1	27/06/2012	Starting version, template (Mark Melenhorst)
V0.5	20/07/2012	First full draft (Mark Melenhorst)
V0.8	24/08/2012	Added missing information (Henk Eertink)
V1.0	31/08/2012	Final version (Mark Melenhorst)

0.3 Document data

Keywords	user test
Editor Address data	Name: Mark Melenhorst Partner: Novay Address: Capitool 15, 7521PL, Enschede Phone: +31 53 485 0438 E-mail: mark.melenhorst@novay.nl
Delivery date	August 31, 2012

0.4 Distribution list

Date	Issue	E-mailer
	Consortium members	al_florence_all@natlab.research.philips.com
	Project Officer	margot.bezzi@ec.europa.eu
	EC Archive	INFSO-ICT-248730@ec.europa.eu

Table of Contents

1 Introduction

In this deliverable the set up of the living lab test is explained, drawing on the results from the controlled homelab environment tests (D6.3). We describe the methodological set up as well as the ethical and safety procedures. In D6.5 the set up of the test will be elaborated into instruments, interview schemes, and survey questions.

1.1 Living lab test

In Florence, end-users have been involved throughout the development process, from focus groups to evaluate early concepts via Wizard-of-Oz testing to a controlled homelab environment test. As a last step within the scope of the Florence project, a living lab test will be conducted. Important lessons can be learnt from that last step: testing the services in real apartments with real users provides us with both technical feedback and with a ecologically valid feedback from end-users.

This end-user test is referred to as a 'living lab test'. Living labs can be understood as an environment for innovation, and as a research methodology (Van Houten et al., 2011). Eriksson et al. (2005), as an example of the latter, define living labs as "a research and development methodology whereby innovations, such as services, products, and application enhancements, are created and validated in collaborative, multi-contextual empirical real-world settings". They highlight the importance of co-creation, development and validation in a real-world setting for the innovation process. Developing and testing in a real-world setting has the advantage of ecological validity: as a researcher you observe the natural behaviour of the target group, provided that data collection is as unobtrusive as possible.

Many data collection methods and techniques exist that help to understand the user's behaviour without disrupting their natural behaviour. Some of them are retrospective in the sense that users are asked to explain about what they did afterwards (interviews, critical incident technique[1], and so on), some of them are concurrent (experience sampling[2]). In addition, often cameras and sensors are used, in which case the user is only disrupted in his natural behaviour to the extent that he or she is aware of the cameras and sensors that collect data about them. In the Florence test we will take an approach that combines the use of cameras and sensors with experience sampling and pre- and post interviews.

Apart from evaluating the added value and potential impact of the Florence robot, we will also address the user experience, as this is in an important aspect of any ICT-innovation in general and for Florence in particular: how do users experience the interaction with the robot?

The 'user experience' refers to "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" (ISO 9241-210). Kavinsky (2010) for instance define user experience as follows:

> *Tthe totality of end-users' perceptions as they interact with a product or service. These perceptions include effectiveness (how good is the result?), efficiency (how fast or cheap is it?), emotional satisfaction (how good does it feel?), and the quality of the relationship with the entity that created the product or service (what expectations does it create for subsequent interactions?)."*

Kavinsky's definition is in line with the traditional usability aspects as defined, among others, by Nielsen (1994): effectiveness, efficiency, learnability, memorability, error Prevention and Recovery, and user satisfaction. Quesenbery (2003) explains these dimensions as follows:

- *Effectiveness:* The completeness and accuracy with which users achieve their goals.
- *Efficiency:* The speed (with accuracy) with which this work can be done.

[1] http://knowledgecentre.openlivinglabs.eu/learn/techniques/critical-incident-technique
[2] http://knowledgecentre.openlivinglabs.eu/learn/techniques/experience-sampling

- *Satisfaction:* How pleasant, satisfying or interesting an interface is to use.
- *Error prevention:* How well the product prevents errors, and helps the user recover from any that do occur.
- *Learnability:* How well the product supports both initial orientation and deeper learning.

However, both in design and in research more and more attention is paid to the affective part of the user experience (referred to as emotional satisfaction by Kavinsky, 2010). This is important for Florence as the user's emotional response to the robot is predicted to be an important determinant of the user's motivation to use or not use the robot.

1.2 Lessons learnt from the controlled controlled homelab environment tests

This section summarizes the lessons learnt from the two controlled homelab environment tests by Philips and Offis. These lessons are the starting point for the set up of the living lab test that is the topic of this deliverable.

1.2.1 Lessons learnt regarding the services

The controlled homelab environment tests have taught us two important lessons:

- Two main categories of application services
- A large number of user interaction remarks

Two main application scenarios:

Users did not clearly divide the services into the classes that we have defined for the controlled homelab environment test. For example, KEETOU and FALHAN are from a user's point of view very similar because both services use video communication and remote control of the robot to assist the elderly. So we will re-cluster the services into two main application services:

- "Communication" service: KEETOU-FALHAN-COLGAM
- "Coaching" service: HOMINT-LIFIMP-AGEREM

Both include safety / security functions. This will make the functionality more clear and easy to understand for the user. In general the backend functionality will not change much but the presentation to the user will be different. In the next chapter the two main scenarios will be described, including the specific research questions we want to have an answer to and what measures we are going to use.

User interaction remarks

The user interface itself also has to be updated to better suit the needs of the elderly users. We received a lot of comments regarding

- Button sizes
- Font sizes
- Speech (loudness, clearness)

These aspects will be improved in the next version of the Florence system that will be evaluated in the living lab test. It is a goal to make those settings as much user configurable as possible.

As another aspect of the 'user interface' the physical structure has to be improved as well. So far the robot doesn't seem to be very stable and is shaking a lot while moving. We will overcome the general stability issues with Pekee by switching to the TurtleBot platform but we have to make sure that the TurtleBot is more stable than the Pekee.

Finally, in the living lab test much attention will be paid to how users experience different interaction modalities (speech, touch, gestures). Which modality or modalities do they prefer? For what reasons?

1.2.2 Input for the living lab test

In regard to the living lab tests in the real homes a series of recommendations can be made. First of all, it will be very important to have a mostly stable system that preferably doesn't need any intervention from a technical supervisor and let the users interact with the Florence System as independently as possible so that real feedback is provided.

The user interface has to be very clear and the functionality of the robot as well. The user shouldn't ask 'what can I do with it?' this should be obvious by design. Navigating through the functions should be self explaining and not much training should be necessary. However, a limited introduction will be part of the test procedure.

If additional hardware is required to be installed in the user's home, these should not need any intervention by the elderly, they should work silently in the background reducing the intrusive nature of the system. Ethical and privacy issues need to be dealt with.

Explanation and introduction to the users should be based on the best way for them to understand and relate each functionality and not on the former names of the service categories.

- "Communication" service: KEETOU-FALHAN-COLGAM
- "Reminder/Notification" service: HOMINT-LIFIMP-AGEREM

Finally, some lessons could be learnt regarding the participants that need to be involved in the living lab test. First, most elderly involved with the homelab test have never tried video chat before and can only compare robotic telepresence to a mobile phone. If possible, it would be good to test with elderly that have already a video conferencing service from a telecare provider (CarintReggeland) such that we can much better compare what the mobility adds to video call experience.

Second in the homelab user tests, most elderly were still physically very fine: they were able to come to Philips and OFFIS to participate in the tests. As such, these elderly slightly fell out of the target group. For the living lab test, it would be good to have elderly participants with some mobility problems and a clear need for help, e.g. elderly that cannot cook anymore, and/or need daily visits from a nurse for medication or other care. This holds especially for the coaching services (e.g. LIFIMP, HOMINT) where the controlled homelab environment tests showed that elderly who were still very healthy did not see the need for coaching applications whereas elderly who were less healthy did see the need. For that reason, participants will be involved that already receive professional homecare.

2 Research questions

In this chapter, we will identify the goals and research questions for the living lab test, given the feedback from end-users we have received and the current state of the technology.

2.1 General research questions

Technical questions:

1. *To what extent is the Florence robot able to autonomously navigate the apartment of the elderly?*

   Navigation in relatively small living rooms is challenging. It is important to identify what the critical situations for navigation are in practice.

2. *To what extent are the elderly using different modalities to reach their goal?*

   The system supports multiple input-modalities (voice, gestures, touch) and output-modalities (graphics, audio, robot actions). What are their preferred input modalities and why?

General user questions.

The Florence living lab test addresses three different groups of users: the elderly, who ultimately have to use the robot, their family members, and professional caretakers. Research questions for the last two groups of users are tied to the specific scenarios that will be addressed in the subsequent section. However, for the first group, the elderly, there are some general research questions that are independent of the scenario that is being tested:

1. *To what extent does the robot offer added-value to the elderly compared to similar non-robot services?*

   The evaluation should provide a clear indication whether users think that the Florence services can best offered by a robot rather than another device (mobile phone, tablet, and so on.) In other words: what is the added value of offering services via a robot?

2. *How do users evaluate the user experience of the robot?*

   The user experience is a general research question, but there are also scenario specific research questions, which are addressed in the subsequent sections. The general questions are:

   a. What improvements to the user experience can be derived from the users' behaviour?
   b. How do users evaluate the user experience?
      i. How do users experience speech recognition?
      ii. How do users experience gesture recognition?
      iii. How do users experience the use of the touch screen?
      iv. Which interaction modality or modalities do users prefer and why?
      v. How do users experience the way the robot moves around in their homes? How close to natural human movements do they perceive the robot to be?
   c. (in particular: the user-robot interaction and the user interface)?
   d. What are the users' affective responses to the robot?

3. *To what extent are users concerned about their privacy? Which design improvements can be derived from their concerns?*
    a. *How do users feel about sensors being installed in their homes? To which extent are they concerned about how the data are used?*
    b. *How do users feel about the camera that is installed in the robot?*
4. *To what extent are users interested in using the Florence services if it were commercially available? What are the conditions and reasons?*

    User acceptance is a precondition for a viable business model for the Florence service, regardless of the pricing model and the stakeholder that will pay for the services. Therefore, a good insight into the reasons and conditions under which users will be interested in using the Florence services on a commercial basis. This information is input for business modelling.

Apart from these general research questions, we want to test a number of application scenarios during the living lab test. Following the lessons learnt from the controlled homelab environment tests, we reclustered the scenarios as follows:

- *The 24/7 remote assistance service*, consisting of KEETOU-FALHAN-COLGAM.
- *A coaching service*, consisting of a reminder and notification service based on LIFIMP. In addition a third class was introduced:
- *Unobtrusive monitoring service*, consisting of sensor-based nactivity detection

In the subsequent sections we will describe the scenarios that will be tested, including the scenario-specific research questions and a set of exemplary metrics. Please note that the scenario descriptions are subject to change, as the technical development is still on-going.

2.2 The 24/7 remote assistance service

24/07 remote assistance is both useful in emergency situations like a fall with professional care providers and just in everyday situations in which family and friends can provide the elderly with practical support.

2.2.1 Description

Fall-handling

The emergency handling service describes a safety functionality of the Florence system. The robot will be used to handle a fall alert. This alert can be detected by different specialized detection systems (which are not developed within the Florence context). The service is designed to increase the actual and perceived safety of the elderly. In the case of a sudden fall, most people are not able to call for help. The robot can act as a mobile monitor and actor and though is able to clarify the situation. The robot can locate the person, try to establish contact and trigger further handling of the situation. Another aspect is the view of the care givers which are able to do a remote check of the fall situation to decrease false alarms. This is done via a telepresence session which is established automatically. So the care giver is able to remotely control the robot and check the environment and the elderly status.

Within the Focus Groups and Wizard of Oz tests, this functionality was appreciated very much by the elderly. Now it is the question if the behaviour of the robot is developed in a right way. The movement of the robot should not be annoying or frightening.

General outline:

Phase 1: Fall has happened, robot handles situation. The elderly is asked to either lie down or sit on the floor (this would be best for a realistic impression but if this is not possible, the elderly can stay seated on a chair). Then the alert should be triggered. This will be done by pressing an emergency button. The robot will then approach the elderly and try to check the situation. Afterwards it will try to establish contact to the user. If this is successful and the user is ok, the alert is deactivated. If the user doesn't react, the robot will call the emergency contact (tele care centre, phase 2).

Phase 2: Tele care centre takes over control. This phase tests the interaction of the tele care personnel with the robot and the user. The robot calls the tele care person. He/she will check the situation and talk to the elderly. During this talk he/she may steer the robot to a different position to get a better view etc.

Keeping in touch

2.2.2

The KEETOU service provides a robotic telepresence service. It's initial purpose was to to increase the social connectedness of the elderly. This is based on the following hypothesis that we want to verify/validate by means of user tests:

- *Robotic telepresence provides an increased feeling of presence with respect to phone and traditional video chat, both for the elderly as well as for the visitor (child, grandchild, good friend, etc.).*

From the user tests of the first implementation, we learned that elderly and family indicated that the main use case they saw was assisting the elderly when needed with all sort of daily issues and problems, like helping with devices. Furthermore, it provided family members with the opportunity to call for a quick chat and actually see whether everything was in order: is the elderly looking good? Is he/she dressed? Is the room cleaned up? Therefore we want to explore this use case further in the second user tests.

Phase 1: Elderly establishes a call.

The elderly can initiate the KEETOU session by first calling the robot with its name (e.g. "Pekee or Max). On hearing its name, the robot will ask "How can I help you", either via the touchscreen or by means of voice. At the same time, if the robot is not close to the person, the robot will also approach the elderly person. However, the robot should be in the same room as the participant. The actual call can be initiated in two ways: via voice or via the touch menu. In case of voice, the elderly can say "Call Anja". The robot will respond with asking for a confirmation: *"Do you want to make a call to Anja?"* the elderly can than answer *"Yes"* or *"No"*. The touch screen will show what the elderly can say. The elderly can also choose to start the session via the touch menu: by first selecting the "video call" option on the touch screen of the Florence robot and subsequently selecting the right person.

Phase 2: During the call, moving around

The elderly test person will go from the living room to the kitchen to show something in the kitchen to her son or daughter. The son or daughter should steer the robot to follow the elderly going to the kitchen.

The goal is to test the "ease of remote navigation" of the robot while communicating with another person, and to find out whether such a scenario is perceived as an advantage over static video chat or video chat via a mobile phone. As a concrete example, the elderly could ask his/her son/daughter how she/he could use the oven functionality of the microwave.

2.2.3 Research questions

Fall-handling

Research questions for the elderly

- How do users evaluate the approach of the robot after the user has fallen?
- How do users evaluate the interaction with the robot?
    - To which extent are the interaction modalities well-adjusted to the physical position of the user (sitting, lying, standing)?

    - Which interaction modality does the user prefer in contacting family members via the Florence robot?
- To which extent does the Florence robot improve the user's feeling of safety compared to a regular alarm button?

Research questions for the caregivers

- To which extent does the caregiver believe that the Florence robot offers added value (in terms of ease of use, efficiency, and effectiveness) to handle emergency situations?
- What is the overall gain in time/effort/money when the robot can provide a first interaction with the elderly?
- How easy is it for care givers to use the remote control for navigation with the Florence robot?
    - To which extent can the Florence robot provide the care giver with auxiliary eyes?
    - How do remote helpers evaluate the control over the orientation of the robot and the way they can point something out to the assisted person?
    - To which extent can it help with false alarms (e.g. user pressed button by accident)?

<u>Keeping in touch</u>

Research questions for the elderly

- Would elderly like to have an video chat solution? With whom and when would they use it (e.g. family, friends, doctor, nurse)? For what purposes can this video chat solution be used?
    - To what extent can remote video presence of family members help the elderly deal with practical things? For what reasons?
    - To what extent does the elderly feel watched by his family member?
    - What are his or her main other concerns?
- Would elderly and their family members appreciate this communication over a traditional phone call? And over video conferencing?
- Does the elderly have a stronger feeling of presence for his family member?

Research questions for the family members

- To what extent does the family member have a stronger feeling of presence for the elderly?
- To what extent does the family member feel more reassured about the well-being of the elderly?
- Is it important that you can move the robot around? That you can follow the elderly? In which situations is this useful?
- What are the main concerns?
- What other applications could it be useful for?

The following technical aspects of the KEETOU are tested:

- *How well does the "assisted remote control" work? How well does it prevent the robot from hitting obstacles during remote control?.*

  Test: A test person will try to drive the robot around and drive the robot into obstacles. It will be tested whether the robot will always prevent being driven into obstacles and how well the robot tries to find a suitable path around an obstacle.

- *What is the latency for the remote control?*

  Test: This will be tested by testing a robotic telepresence session over a large distance (e.g. 50 kilometres) and not only in the lab. The roundtrip latency of a robot steering command will be measured.

- *How is the quality of the sound perceived, especially with respect to echo cancellation and background noise?*
  Test: The maximum distance will be tested of how far the person at the robot can be while still being understandable to the other side. This will be done in different conditions.

2.2.4 Exemplary metrics

For each of the research questions metrics need to be defined and instruments need to be found or developed. For fall handling and keeping in touch examples of metrics could be:

<u>Fall-handling</u>

- Perceived level of safety compared to a regular alarm button
- Perceived user experience (Nielsen, 1994)
  - Effectiveness
  - Efficiency
  - Learnability
  - Memorability
  - Error Prevention and Recovery
  - User satisfaction (including emotional satisfaction)
- Perceived added value compared to regular alarm button
- Perceived trust and privacy with regard to the caretaker who can see the elderly through the robot's camera

<u>Keeping in touch</u>

- Social connectedness (Lee & Robbins, 1995), in comparison to using other devices (smartphones, tablets, and so on)
- Awareness, in comparison to using other devices (smartphones, tablets, and so on)
- Perceived added value compared to regular phone calls or Skype connections
- Perceived trust and privacy with regard to the caretaker who can see the elderly through the robot's camera
- Perceived user experience (Nielsen, 1994)
  - Effectiveness
  - Efficiency
  - Learnability
  - Memorability
  - Error Prevention and Recovery

- User satisfaction (including emotional satisfaction)

2.3 Coaching

The second class of services to be tested is coaching. In the homelab environment tests we have found out that elderly benefit from coaching in the form of reminders for things they need to do at fixed times (such as taking their medication, or blood pressure, or certain appointments). In the homelab environment test this was referred to as the LIFIMP-scenario. In the living lab test blood pressure measurement, physical exercise, and weight measurement are tested.

2.3.1 Description

Blood pressure measurement

Within the blood pressure module, the goal is to measure the blood pressure of elderly. This scenario is designed for elderly that have problems with their blood pressure and have to measure their blood pressure on doctors' orders several times a day. Florence can assist the elderly in reminding them to measure the blood pressure and guide them through the process. The reminder can be entered in the agenda by the user himself or by the doctor.

Florence will help the user to measure the blood pressure by guiding him or her towards the meter, explaining how to use it and give feedback on the results. The data is stored, and the results are shown to the user. No historical data is used. The researcher will explain to the user that no medical consequence whatsoever should be based on the measurements (see the informed consent form in Appendix A).

If a test participant does not need to take his blood pressure for medical reasons, he or she is asked to do anyway, because participants are likely to be able to experience what it is useful for and what added value the robot can offer.

Physical exercise

In this part of the LIFIMP scenario, the user is encouraged to walk outside. The Florence robot will ask the user how his/her energy level is, and what activity they would like to do (biking of walking). Florence will suggest a route for the user to walk/bike based on the current position of the user (location of the house). Florence will then show a map with the suggested route. There is no feedback-loop that determines whether the user has actually completed the activity.

After watching the television for an hour, Florence will come to the user and will suggest to do some exercise. Florence will then go back to its original position. Please note that some work will have to be done on this scenario before it can be tested. The scenario has to be updated and the ability to use activity recognition as a trigger has to be implemented.

Weight measurement

In weight measurement, the user is able to measure their weight and get an overview of the progress made throughout the last measurements. Even though the activity of measuring ones weight is something the user will not do on a daily basis, we will include it in the user test. An item will be scheduled in the agenda and a reminder will be given to the user.

2.3.2 Research questions

Blood pressure

- What are the benefits for using a mobile, interactive robot to help the elderly measuring their blood pressure?

- Does the compliance of elderly to measuring their blood pressure improve by using a mobile, interactive robot to remind them and help them?

Physical exercise

- To what extent do the elderly feel more motivated to exercise when a robot reminds them to take a walk? What is the added value of a robot delivering this reminder?
- How do users feel about the use of the number of hours they are watching tv as the basis for the timing of the reminder?
- To what extent do the users feel watched?

Weight measurement

- *What is the added value of weight measurement via the Florence robot?*
- *How do users feel about reminders to measure one's weight?*
- *How do users feel about their weight measurements being stored, in terms of privacy?* Even though this is not the case in the user test, in a more market-ready version of Florence, this will be possible. How do users feel about their data being stored and used for weight coaching purposes?

2.3.3 Exemplary metrics

Below we define some exemplary metrics that may be used in the user test:

Blood pressure

- Compliance: the extent to which the participant actually takes his or her blood pressure
- Perceived added value compared to other types of reminders (their spouse, calendar on a cell phone, and so on)

Physical exercise

- Compliance: the extent to which the participant actually takes a walk
- Attitude towards physical exercise

Weight measurement

- Compliance: the extent to which the participant actually takes his or her weight
- Perceived added value compared to other types of reminders (their spouse, calendar on a cell phone, and so on)

For all of the above:

- Perceived user experience (Nielsen, 1994)
  - Effectiveness
  - Efficiency
  - Learnability
  - Memorability
  - Error Prevention and Recovery
  - User satisfaction (including emotional satisfaction)

2.4 Activity detection

2.4.1 Description

This class of services was introduced to evaluate whether activities can be detected based on sensor data and what users perceive as the added value of this activity detection. Activity detection is often used to monitor elderly in nursing homes. However, in this test we want to investigate what activity detection can contribute when people still live ate home.

We distinguish two type of activities:

- Hard activities: those activities that can be directly derived from the values sensors provide. Such as: opening a drawer, switching on the tv, or walking into a room.
- Soft activities: those activities for which reasoning is necessary to detect them: such as 'making lunch'.

Below we list the activities that we could be detected at this point:

- Breakfast
- Lunch
- Dinner
- Visitor
- Entertaiment
- Empty house
- Sleep
- Hobby
- Personal hygiene
- Hot beverage

For these activities the following sensors are used:

- PIR-sensors
- Contact sensors (windows, fridge, cupboard, front door)
- Power sensors (Computer, DVD, tv, microwave, coffee machine, and ventilation)
- Door bell
- Temperature (measured in bedroom)
- Light sensor

The precise sensor configuration and the activities that can be detected need to be determined once we know the layout and the configuration of the apartments.

2.4.2 Research questions

The following research questions will be investigated:

1. *What type of soft and hard activities can be detected in a short time frame in a natural setting?*

   The system is used only for a few days. This means that it will not be possible to detect changes in normal activities of daily life. However, it is likely that easy-to-detect activities still provide information about the normal way of living in the apartments.

2. To what extent are users aware that sensors are installed in their apartment? To what extent does this awareness change their behaviour?
3. What is the perceived added value of activity detection? What use cases do participations see? To what extent are users concerned about their privacy in these use cases?

2.4.3 Exemplary metrics

- Number of activities correctly detected
- Number of erroneous activities detected
- Attitude towards sensor-based monitoring
- Perceived added value
- Perceived level of privacy

3 Living lab methodology

This chapter provides the methodological outline for the living lab test. This outline is the basis for the development of detailed procedures, and instruments (metrics, interview and observation schemes) over the course of the next months, which will be reported in D6.5.

3.1 Participants

For this user test, a number of different target groups will be involved. The selection criteria for each of these target groups are discussed in 3.1.1. We are collaborating with CarintReggeland to find the participants we are looking for in the user test. The background of this collaboration is discussed in 3.1.2. Finally, the recruitment procedures are explained in 3.1.3.

3.1.1 Target groups

The following target groups will be involved:

- *Elderly: the users of the Florence services.* (n=5)

  Selection criteria

  - Participants should live alone in their own home: the current implementation of the Florence robot and associated sensors cannot deal with households with more than one person.
  - Participants should not have serious physical or mental barriers, but should be in need of some support. The Florence services are targeted at people that have some need of health and well-being support without serious physical or mental barriers. We therefore target our efforts at participants that receive some care at home via CarintReggeland, as these clients best meet the aforementioned criteria.
  - Participants should not have pets, because it disrupts both the navigation of the robot and the activity detection sensors.
  - Participants should preferably use medication. This is of course desirable for the medication reminder. However, it is not strictly necessary since using medication is something people can imagine. They are likely to have experience with it, even when they do not take medication at the time of the test. In the user test, we can use fake medication reminders that do trigger robotic actions.
  - Men and women should be equally represented in the sample, if possible.

- *Family members and informal caretakers* (n=5)

  Family members are selected and recruited via the elderly participants. This could for instance be a child or grandchild.

  - Participants should be available at a certain time to test the awareness-related scenario's.
  - Participants should be able to connect with their family members via a Skype connection .
  - Participants are preferred who cannot visit their elderly as often as they would like. This criterion was inserted to involve participants that would benefit from the KEETOU service most.

- *Healthcare professionals*

  Health care professionals should be recruited to test the FALHAN-scenario: a health care professional should navigate and control the robot and get in touch with the elderly to handle the emergency situation. The following criteria apply:

  - The healthcare professionals should be available at the time of the test
  - The healthcare professionals should be able and willing to learn how to operate the robot

- The healthcare professionals should preferably be experienced in remote fall handling (for instance, via a regular phone line, or via a fixed video system)

The number of different professionals should be discussed with CarintReggeland. Depending on their preference and practical feasibility, either one professional for one participant might be recruited, or multiple participants may be assisted by the same professional.

- *Innovation managers* (n=2)

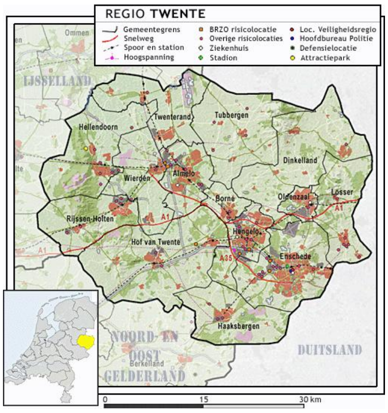
  Innovation managers will be recruited for an interview to explore the business potential of the Florence services.

3.1.2 Collaboration with CarintReggeland

CarintReggeland[3] is a regional care provider in The Netherlands, located in the region Twente, headquartered in Hengelo. Novay has worked with and for CarintReggeland. The latest joint effort was the user-test of the TalkMeHome[4] service (2011).

Some information about CarintReggeland Groep[5]:

Metric	2010	2011
#intramural customers	1.809	1.723
#extramural clustomers	4.596	4.514
#beds/intramural	1.872	1.723
Employees	**2010**	**2011**
# employees on payroll	5.042	5.170
#FTE 's (avg)	2235,9	2.248,3



CarintReggeland operates over 100 geographical locations in the region (offices, care centers, daycare, …). They also operate a tele care center, equipped with the Verklizan UMO product[6].

3.1.3 Recruitment procedure

The following steps are taken to recruit participants:

1. In August, CarintReggeland will distribute a one-page flyer to the residents of approximately 5 apartments.
2. In September, they will call potential participants and ask whether they want to take part in the study, aiming at approximately 10 apartments. If more than 10 people are willing to participate, the participants will be selected by Novay and CarintReggeland, based e.g. on the layout of their apartment. The robot needs at least 50 cm navigation-space around furniture; it is likely that not all apartments will be suitable for the pilot.

[3] www.carintreggeland.nl
[4] www.talkmehome.nl (in dutch)
[5] Taken from: http://carintreggeland.nl/dbimages/bijlagen/Maatschappelijk%20verslag%202011%20vastgestelde%20versie%2017%20april%202011.pdf (in dutch)
[6] http://www.verklizan.com/index.php?id=3&L=0

3. Early October, a user researcher and an engineer will visit the 10 participants . During this visit the study is explained to the participants and questions are asked to determine whether a potential participant meets the selection criteria. Furthermore, their apartment will be evaluated against the requirements (see next section).
4. By the end of October the final selection of approximately five participants will be made.

3.2 Test environment

An undisrupted use of the robot requires that it can navigate freely across the apartments it will be tested in. Therefore, during the selection of participants a number of requirements to their apartments should be considered. It is necessary that the apartments:

- have a broadband internet connection (necessary for the awareness scenario)
- have sufficient[7] room between pieces of furniture to allow for unobstructed navigation
- have a tile floor, wooden floor, or a non-deep-pile carpet
- have no doorsteps or small doorsteps, to avoid disrupting the robot's movements

As part of the recruitment and selection procedure the researchers will visit the homes of the participants to evaluate whether their apartments will meet the selection criteria. This is an important step, not only for technical reasons but also for the safety of the participants.

3.3 Procedure

3.3.1 General outline

1. Introduction (day 1 – morning)
2. Installation of sensors and cameras (day 1 – morning)
3. User-evaluation of the services (day 1, 2, to be discussed with participants; total test time approximately one day)

   As part of the evaluation of the keeping in touch service the second target group, the informal caretakers, will be interviewed.
4. Post-interview (day 2)

   After the evaluation, the elderly will be interviewed to discuss their experiences with the Florence services.

3.3.2 Introduction

The introductions serves the following purposes:

1. *Selection of participants that have expressed their interest*
2. *Introducing the evaluation*

   In this step, the study is introduced. Participants are explained what will happen, what is expected from them, what data is collected and how the data will be used. Practical implications are discussed, including the sensors that will be installed and any changes that might need to be made to the position of the furniture to allow for easy navigation.

   In this part particular attention should be paid to safety concerns regarding the use of the robot. In Deliverable 5.2, section 9 risks have been identified and measures have been suggested.
3. *Signing informed consent forms*

   Once the participant has fully understood what participating in this study means, they are asked to explain the test in their own words. When they have understood everything, they are requested to sign an informed consent form.

[7] What is sufficient, is determined by the new hardware for the robot that has just been received.

4. *Introducing the robot service*

   The robot service is explained to the participants. A demo script will be prepared to precisely determine the contents of the explanation. The demo should provide enough information to enable the participants to use the robot, but should be as neutral as possible to avoid disrupting the natural behaviour of the user.

3.3.3 User evaluation

Over the course of the next few months, the procedures for each of the services need to be refined. The procedures will depend on the technical state of the services and the extent to which we can observe the user via cameras(see section 3.4). However, some service-specific procedures can be mentioned already.

<u>Fall handling</u>

Obviously, the evaluation of this service has to be simulated. The user will be provided with an alarm button. An item will be put in the agenda that asks them to walk to a random point in the apartment and push the alarm button. The user can choose to sit down or lie down, depending on their health. The timing needs to be discussed with the care giver that will handle this 'emergency situation'.

<u>Keeping in touch</u>

A time-slot needs to be scheduled at which the user can call a family member. This has to be discussed before the user test, since both the elderly and his or her family member should be present. A reminder 15 minutes prior to the scheduled time may be scheduled in the agenda. If the elderly test participant has no Skype account, a Skype-connection should be made in advance between the family members of the elderly participants and the elderly participant via a 'general' Skype name.

<u>Agenda reminders</u>

The user's agenda must be filled with robot-supported activities. Potential activities:

- triggering an outside walk
- taking medication
- measuring the weight
- measuring blood pressure
- normal, private, reminders

The standard behaviour is that the robot moves to the user and the user interacts with the robotic service using his/her preferred modality. The items put in the agenda will cause the robot to deliver a reminder about the task to the user.

<u>Activity recognition</u>

Due to the short duration of the user test with a single elderly person, it is not guaranteed that multiple activities can be recognized. However, we will deploy sensors in the home that make it possible to determine what type of activity is being executed by the elderly. This service is not visible to the user; the recognized activities are stored in the Context-Mgt Service, and cannot be retrieved by the elderly him- or herself. The sensors will remain in place throughout the user test.

3.4 Data collection

3.4.1 Pre-interview with the elderly

The semi-structured pre-interview is combined with the introduction. The interview-part of the introduction starts after the consent forms have been signed and the demo of the Florence scenarios has been provided. Questions will be asked about their background (ICT-experience, health situation, social situation). Furthermore, we will ask them for an initial evaluation of the Florence robot and its services. Topics that will be addressed are:

- The estimated usefulness of the services
- The estimated added value of the services compared to alternatives
- The estimated self-efficacy related to the use of the robot: to which extent do the participants feel that they are capable of using and operating the robot and its services?

An interview scheme will be constructed that will address these topics. The interview will include the quantitative metrics that were introduced in the previous Section 3.

3.4.2 Observation and logging

To allow for an unobtrusive observation of the participant's behaviour we will use video cameras combined with event logging. A coding scheme has been developed that will track specific behaviours. The level of detail with which the user can be observed depends on the camera installation. This will be investigated in September.

Emotions	Record of participants voicing and affective response to using the robot or its services
Verbal expressions	Record of participants voicing their thoughts about the robot, its services, user experience, usefulness or added value
User-robot interactions	Observations of participants dealing with the robot. This can be (succeeded) attempts to control the operation of the robot by means of gestures, voice, or by using the touch screen
Technical issues	Observations of the user or the robot experiencing problems as the result of technical malfunctioning

3.4.3 Experience sampling

Right after a service has been tested, a brief survey with closed questions will be administered via the touch screen in order to get reliable in-situ feedback. Questions may comprise the user's emotional response to the service, a first holistic evaluation, or any other service-specific metric. For instance, in keeping in touch, one could ask how close the participant feels to the family member. The questions will be service-specific to avoid learning effects.

3.4.4 Post-interview

The semi-structured post-interview consists of the following parts:

1. *Critical incidents*: the respondent is asked for critical incidents: particular positive or negative experience that were meaningful to the respondents and why this was meaningful to them. This will yield their general impression, as well as those aspects of Florence that were most important to the participants.

2. *Critical incidents from observation and logging*: observation and logging will produce particular critical incidents. The recordings of the incidents will be shown to the respondent and then discussed with them, provided this is possible with the camera set up. If not, the notes from observation will be used to select critical incidents.
3. *Evaluation of the scenarios*, starting with a questions about what scenario is most and least appealing, questions about how they evaluate the current combination of scenarios, and the respondent's ideas about scenarios they would like to add. In this part we will also address the differences between what they have experienced and their expectations beforehand (based on the introduction).
4. *Business potential:* questions about whether the respondent would like to have the service if it were commercially available, including questions that ask for their willingness to pay for it.

With respect to part one and two, the incidents will be discussed to elicit feedback regarding the most important user experience aspects (Nielsen, 1994): learnability, ease of use, fun of use, comprehension, satisfaction, and the user's affective response to the robot.

Furthermore, the added value of a robot will be addressed: given the service, what added value does the robot offer compared to other implementation? For instance, do participants feel less connected to a family member when they call them via Skype on a laptop compared to a robot?

Finally, the participants is asked for suggestions for improvements to the robot or its services.

3.4.5 Reflection

Throughout the user test the researchers will keep a log of occasions in which the living lab test proves to yield observations or design recommendations that could not be derived from a homelab test in a controlled environment. This concerns both user behaviour and the behaviour of the robot in a 'real' apartment.

3.5 Ethics and safety

Ethical procedures are important when testing in a real-life environment with end-users. CarintReggeland does not have formal procedures in place for non-medical research such as the current study. However, the study was checked and approved by the 'Medisch eintverwantwoordeiljke' (manager responsible for medical issues) of the CarintReggelandgroep, Henk Snijders and approved by Annemarie Asbreuk, director Additional Services (see Appendix A).

3.5.1 Informed consent

Informed consent is the most important ethical principle related to any user research. Therefore, we have prepared a form that should be signed by every participant only if they have understood everything that is mentioned. A draft version of this form is displayed in Appendix B (in Dutch). The form was checked, revised, and then approved by Henk Snijders, 'Medisch eintverwantwoordeiljke'.

The form stresses:

- What is expected from the participants (as explained in this chapter)
- That participating is completely voluntary; that participants are free to stop their participation at any moment, without losing their rights on (monetary) compensation for their participation
- That the research data is only used in a manner that cannot be traced back to the individual participant.
- That video recordings can be used for publications or presentations, but only after making the participant unrecognizable.
- That the connection between the participant's private data (name, address, and so on) is destroyed

3.5.2 Safety

In Deliverable 5.2 a number of risks have been identified. In this section we will report on how we want to deal with these risks to prevent the Florence robot from causing harm to the participants. The safety management is discussed here for the sake of the participants in the user test. The general safety issues for the final services have been discussed in Deliverable 5.2. These risks will be explained to the participants as well, as part of the informed consent procedure.

- **(General) Robot causes elderly to fall**

  The robot as an autonomous mobile device might bring about the risk of causing the elderly to fall.

  It is important to implement remote control functionality to navigate the robot. The main goal of the user-test is to evaluate the acceptance of services on an autonomous mobile robot. If the navigation is not properly functioning (e.g. due to limited space between furniture), it will strongly influence the perception of the user towards service robots. During the user-test it is therefore recommended to have a live video-image of the living room on the researcher's laptop in order to be able to take over the robot-navigation. An emergency button will be implemented to stop all activity if necessary. The researcher can also manually take over the navigation of the robot in case of emergency.

- **(General) Robot falls over**

  Robot falls over. This could lead to robot forming an obstacle to the user and potentially cause the elderly to fall.

  An important criterion for the selection of the apartments is that the robot can navigate freely without any obstacles that could be dangerous. An emergency button will be implemented that can be controlled both by the user and by the researcher. If anything is about to happen, the researcher can push the button, which will result in an immediate stop of any movement. The researcher can also manually take over the navigation of the robot in case of emergency.

- **(General) Unauthorized access**

  An unauthorized person remotely controls the robot, without the consent of the user. Note that because of their potential household robots might become an interesting target for "hackers".

  This is not the aim of the user-test. It is not yet a commercial product.

- **(General) Robot creates an electrical short circuit**

  The robot is an electrical device and as such there might be the risk is of an electrical short circuit. This could either happen at the docking station or at the robot. An electrical short circuit could create an electrical shock for the user and even create fire.

  A fire-extinquisher will be purchased for the apartment at the cost of about 60€.

- **(General) Robot blocks important passage ways**

  Robot blocks important passage ways of the elderly. This could be doorways, blocking access to cupboards, couch etc.

  This can be handled using the emergency-remote control functionality.

- **(General) Robot falls of the stairs**

  Robot does not detect the stairs and drives down the stairs causing it to fall over.

  This risk is not applicable to the current set up of the living lab test, as only participants that live in an apartment (without stairs) will be included.

- **(General) Robot behaves erratically**

  Robot malfunctions and starts behaving erratically resulting in e.g.:

  - Making loud sounds

  - Driving around in a uncontrolled and chaotic way

  Emergency button and remote control should take care of unexpected behaviour that is the result of technical malfunctioning.

- **(General) Robot startles elderly**

  Robot might startle elderly because of sudden movements, unexpected presence (e.g. robot "sneaks" behind elderly)

  This is unlikely to happen, since the robot is noisy enough to announce its arrival. The explanation of the test beforehand should be clear enough to avoid unexpected behaviour. Emergency button and remote control should take care of unexpected behaviour that is the result of technical malfunctioning.

- **(LOGSYS) The displayed health recommendations are wrong**

  The displayed health recommendations are wrong

  This risk is covered by two measures:

  1. No real historical health data is used. The displayed data is made up. This will be explained to the participants during the introduction.
  2. No explicit health recommendations are provided. Medication reminders should be scheduled by the participants manually. The logged health data (blood pressure, weight) is presented without advice.

     The only advice that might be classified as health recommendations is that when participants watch tv for more than one hour, a suggestion is offered to do something else - like taking a walk.

- **(AGEREM) Incorrect Posology Information**

  The user gets a reminder to take their medication but is provided with the wrong information (e.g. incorrect medication or dosage)

  During the user test the user will not be asked to take 'real' medication, but a substitute (like peppermint drop). The informed consent form explains that the user test does not have any medical relevance with regard to medication use, medication reminders, or measurements.

- **(AGEREM) Failure to deliver a medication reminder**

  The user has become dependent on the AGEREM service when remembering to take medication. The user should have received a reminder to take medication but didn't. This may be due to the fact that the robot has a low battery or a system malfunction.

  This is unlikely to happen since the user test lasts at the most only two days.

- *(***FALHAN). Robot can't connect to Tele care Service**

  A fall has been detected (alarm-button pressed). The robot cannot contact the tele care center.
  The robot must call the tele care center via standard telephony.

- **(FALHAN) Panic button is activated, the robot doesn't receive the signal**

  The user has fallen and used the panic button. Due to connection problems or interference the robot does not recognize the signal.

  During the living lab test the fall handling scenario is not tested for real, but it is simulated. Instead of falling, participants can either sit down or lie down somewhere, on their own initiative. This makes the risk of a malfunctioning tele care service not dangerous anymore.

- **(FALHAN) The robot can't reach the person**

  The user has fallen and used the panic button. Due to a closed door or an obstacle, the robot can't reach the person.

  The researcher is present in the apartment. When assistance is required and the robot cannot reach the participant, the researcher can call for assistance. Please note that the FALHAN-scenario is simulated, meaning that this situation will only occur if a participant accidentally falls.

3.6 Data analysis

Data will be analyzed with the purpose of deriving the following types of information

- A list of improvements. Improvements include:
  - the services
  - the user-robot interactions (gestures, user interface of the touchscreen, speech)
  - autonomous behaviour of the robot (navigation, sensing, activity detection)
- A list of lessons learnt with respect to the user-centered design cycle

  The step from a homelab test in a relatively controlled environment to a living lab test in 'real' homes is quite substantial. The living lab test will result in new lessons learnt that could not be learnt from the homelabtest. These lessons will be recorded throughout the whole evaluation process.

4 References

Eriksson, M., Niitamo V.-P- and Kulkki, S. (2005): "*State-of-the-art in utilizing Living Labs approach to user-centric ICT innovation - a European approach*". Center for Distance-spanning Technology. Lulea University of Technology, Sweden. Nokia Oy, Centre for Knowledge and Innovation Research at Helsinki School of Economics, Finland. Retrieved from: http://www.vinnova.se/upload/dokument/Verksamhet/TITA/Stateoftheart_LivingLabs_Eriksson2005.pdf

Kuniavsky, M. (2010). *Smart Things: Ubiquitous Computing User Experience Design: Ubiquitous Computing User Experience Design.* Burllington, MA: Morgan Kaufmann.

Lee, R.M. & Robbins, S.B. (1995). Measuring belongingness: The Social Connectedness and the Social Assurance scales. Journal of Counseling Psychology, Vol 42(2), p. 232-241

Nielsen, J. (1994). Usability Engineering. San Francisco, CA: Morgan Kaufmann Publishers.

Appendix A Ethical approval by CarintReggeland

Novay
T.a.v. de heer Eertink

Postbus 589
7500 AN Enschede

contactgegevens

T 074 367 70 55

www.carintreggeland.n

postadres
Postbus 506
7550 AM Hengelo

bezoekadres
Boortorenweg 20
Hengelo

KvK 41027001

kenmerk	AA/hs
datum	30 augustus 2012
telefoon	074-367 7197

Geachte heer Eertink,

Het Florence-project heeft een onderzoek opgezet om de toegevoegde waarde van een robot en van de voor deze robot ontwikkelde diensten voor het welzijn van ouderen te evalueren.
Op uw verzoek is de opzet van dit onderzoek getoetst aan ethische richtlijnen. Omdat het geen (para)medisch onderzoek betreft, maar een onderzoek naar welzijnsdiensten, zijn de medisch-ethische toetsingsregels niet van toepassing. Wel zal de cliënt die meewerkt aan dit evaluatieve onderzoek van tevoren schriftelijk moeten aangeven dat hij/zij volledig akkoord is volgens het principe informed consent. Verder zal de cliëntenraad (Thuiszorg) ook over dit onderwerp geïnformeerd worden. Het bewijs dat de cliënt volledig akkoord is zal, alvorens het onderzoek start, getoetst worden door de medisch eindverantwoordelijke van Carintreggeland.

Hierbij laten we u weten dat Carintreggeland geen ethische bezwaren heeft tegen het *living lab*-gebruikersonderzoek dat in het kader van het EU-project *Florence* (ICT-2009-248730) wordt uitgevoerd en waarvan Stichting Novay een van de partners is. Carintreggeland is akkoord met de opzet van het onderzoek, de procedure voor geïnformeerde toestemming van de deelnemers, de beperking van veiligheidsrisico's, de dataverzameling en de bescherming van de privacy, zoals dit in het rapport *'Report on the Testing and Evaluation Methodology for the Living Lab Testing'* is verwoord.

Ik hoop u hierbij voldoende te hebben geïnformeerd.

Met vriendelijke groet,

Drs. A. Asbreuk

Appendix B Informed Consent Form

Ik, ________________________________ (*naam participant*)

stem toe mee te doen aan een onderzoek dat uitgevoerd wordt door ***Novay*** en uitgevoerd in samenwerking met de Carint-Reggeland Groep.

Onderzoeker namens Novay is:______ _________________________ *(naam)*

Ik ben me ervan bewust dat deelname aan dit onderzoek geheel vrijwillig is. Ik kan mijn medewerking op elk tijdstip stopzetten en de gegevens verkregen uit dit onderzoek terugkrijgen, laten verwijderen uit de database, of laten vernietigen.

De volgende punten zijn aan mij uitgelegd:

1. Het doel van dit onderzoek is het onderzoeken van de mening van eindgebruikers (ouderen en zorgprofessionals) over het *nut en gebruiksgemak* van **robots voor welzijn thuis**. De gegevens die verkregen worden uit dit onderzoek zullen worden gebruikt om de functies van de robot te verbeteren.
2. Het onderzoek vindt plaats in het appartement van de bewoner. Het onderzoek bestaat uit verschillende onderdelen:
   - Een intakegesprek tussen de onderzoeker en de bewoner waarin de gang van zaken uitgelegd wordt. Dit wordt enkele weken voor het onderzoek gepland.
   - De dag voor het onderzoek worden eerst enkele sensoren geïnstalleerd in het appartement (zonder schade te veroorzaken). Hiervoor maakt de onderzoeker een afspraak met de bewoner.
   - Daarna volgt het eigenlijke onderzoek waarin de bewoner met de robot werkt. Dit wordt in overleg gepland; normaal gesproken in 2 dagdelen van ± 2 uur.
3. Er behoort geen stress of ongemak voort te vloeien uit deelname aan dit onderzoek.
4. De gegevens verkregen uit dit onderzoek zullen anoniem verwerkt worden en kunnen daarom niet bekend gemaakt worden op een individueel identificeerbare manier.
5. De onderzoeker zal alle verdere vragen over dit onderzoek beantwoorden, nu of gedurende het verdere verloop van het onderzoek.

Handtekening onderzoeker: …………………………………… Datum: …………………..

Handtekening participant: …………………………………… Datum: …………………..

Toetsing Medisch eindverantwoordelijke Carint ReggelandGroep:…………… Datum……………..