ICT-2009.7.1

KSERA Project

2010-248085

Deliverable D5.3

Formative Evaluation

31 January 2012
Public Document

©KSERA consortium

Project acronym:	KSERA
Project full title:	Knowledgeable SErvice Robots for Aging
Work package:	5
Document number:	D5.3
Document title:	Formative Evaluation
Version:	1
Delivery date:	31 January 2012 (month 24)
Actual publication date:	30 January 2012
Dissemination level:	Public
Nature:	Report
Editor(s) / lead beneficiary:	Johannes Oberzaucher (CRT) Franz Werner (CRT)
Author(s):	Johannes Oberzaucher (CRT) Franz Werner (CRT) Hadas Lewy (MAC) John Lemberger (MAC) Katharina Werner (CRT)
Reviewer(s):	Cornelius Weber (UH) Luca Feletti (Consoft)

©KSERA consortium

Glossary of terms used

Acronym/terms	Definition
ADL	Activities of Daily Living
AR	Assistive Robotics
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case report form
ER	Emergency Room
FEV1	Forced expiratory volume in 1 second
FVC	Forced expiratory vital capacity
FMECA	Failure Modes Evaluation and Effects and Critically Analysis
GOLD	The Global Initiative for Chronic Obstructive Lung Disease
HCI	Human Computer Interaction
HRI	Human Robot Interaction
iADL	Instrumental Activities of Daily Living
ICF	International Classification of Functioning
KSERA	Knowledgeable Service Robots for Aging
KSERA-system	A smart home environment and a mobile platform, i.e., Nao
KSERA ubiquitous home	A smart home environment
Nao	Humanoid Robot
UCD	User-Centered Design
SAR	Service Assistive Robot
Scenario	Usage narrative
Test case	Scenario based sequence for evaluation purposes
TOT	Task oriented training
Use case	Describe the functional requirements
WHO-ICF	See International Classification of Functioning
WP	Work package
User	Primary User of KSERA system: (a) persons with COPD and (b) older
PT1	Prototype 1
PT2	Prototype 2
TP	Trial participant
TC	Testcase

Contents

Glossary of terms used 4
Contents 5
Executive summary 7
Purpose of this deliverable 7
Suggested Readers 7
Relationship to other documents 7
1. Introduction 8
1.1. Main target of the formative evaluation process 8
1.2. Evaluation model and strategy 8
1.2.1. System performance 9
1.2.2. User acceptance, Perception of the social impact and inference to impact on quality ... 9
1.2.3. User experience concerning the human perception of the robot (HPR) 11
2. Final evaluation test-setting 12
2.1. Description of Test Users 12
2.1.1. Trials 12
2.2. Test setting 12
2.3. Test procedure 16
2.3.1. Preparations prior to the test 16
2.3.2. Reception of the user 16
2.3.3. Pre Test Phase - initial questionnaires 16
2.3.4. Explanation of the KSERA-system 16
2.3.5. Demonstration of test-cases 16
2.3.6. Post test phase - final interviews 17
2.3.7. Discharge and planning of future tests 17
2.4. Limitations concerning the analysis results 17
2.5. Changes during the test phase 18
3. Results – evaluation report 18
3.1. Test-statistics 18
3.2. Results of the Technical Performance Evaluation 18
3.2.1. Localization, Navigation and Movement 19
3.2.2. Performance of Interaction and communication metrics 22

3.3. Performance of data gathering metrics 25
3.3.1. Evaluation of the duration of tasks 25
3.4. Summary of the technical statistical evaluation 27
3.5. KSERA QoL Scale - Usability Evaluation 29
3.5.1. General impression of the system 29
3.5.2. Emotional feeling concerning the system usage, test flow and presented scenarios 30
3.5.3. Level of satisfaction concerning the system functionalities and performance ... 31
3.5.4. Physical measurement performance 32
3.5.5. Training support 33
3.6. Results of the Human Robot Interaction – questionnaire survey 36
3.6.1. KSERA HRI scale – characteristics of the robot 36
3.6.2. KSERA HRI scale – attitude towards the robot 37
3.6.3. Godspeed Questionnaire evaluation 40
3.7. Summary KSERA QoL abilities and HRI evaluation 41
3.8. Comparison between robot system and touch screen solution 43
4. Summary and Conclusions 46
5. Literature references 50
6. Annex – Qualitative user comments 52
6.1 Austria, 1st iteration 52
6.2 Austria, 2nd iteration 55

Executive summary

This document describes the Formative evaluation trials (procedure, results, findings and discussion) within the project KSERA. The trials were conducted in real life conditions in senior citizen homes in Austria and Israel together with a number (n=16) of older (age = 70-91) trial participants.

The test system (KSERA prototype 1) consisted of a humanoid Social Assistive Robot (SAR) system, together with a ubiquitous sensor system and a smart home solution (see KSERA deliverable D5.2). The trials were performed in several iterations and were observed by members of the KSERA research team.

During the trials a strong focus on the technical performance was set. Besides that also usability, Human-Robot-Interaction and quality of life parameters were evaluated.

The results verify the KSERA approach and offer the opportunity for enhancements and further design decisions for the KSERA prototype 2. These results and lessons learned are condensed in conclusion tables at the end of the evaluation section in chapter 3.

Purpose of this deliverable

The deliverable D5.3 **Formative Evaluation** describes the formative evaluation process, methods and results of the KSERA system trials.

Suggested Readers

This deliverable is intended for the general public. It can be read by those willing to know more about the use of robotics systems and their applications in Ambient Assisted Living and shall give a deep insight of the methods and metrics used during the formative evaluation and the results gathered.

Relationship to other documents

The trial description is strongly connected to the trial plan document D5.1 and the first prototype D5.2. The metrics used during the trials are taken from the public deliverable D1.3 “SAR metrics” where they are explained in further detail.

1. Introduction

1.1. Main target of the formative evaluation process

The main target of the formative evaluation process was to verify the KSERA prototype 1 in terms of technical functionality and reliability as well as in aspects of usability, quality of life and human robot interaction. These results form the basis for the development process of PT2, which can be seen as a continuous loop of collecting data, analyzing the shortcomings, making recommendations for improvements, redesign, development and reevaluating.

Due to the rapid evaluation process and rich and explicit data expected, the impact on the future system design was considered to be high. For this reason the evaluation was started already at an early stage during the system development at different labs of partners and concluded with a final user trial. The final formative evaluation, including the user trial, took place at two sites: Austria and Israel. The formative evaluation of the first prototype will guide the development and research in the second iterative design stage of the project.

The evaluation metrics described in the KSERA WP1 deliverable D1.3 SAR metrics (public document) are used in a real user environment. Special attention was directed towards personal safety of end users involved in the tests.

The results of the formative evaluation are used in KSERA WP1-WP2-WP3 and WP4 to make further refinements to the research questions and algorithm development. The ethical advisory board advised and guided the trials, supporting the work during the preparation and trials with a report on ethical topics related to the trial sites and trial procedures.

1.2. Evaluation model and strategy

The main aim of the Formative Evaluation was to test the first KSERA prototype in a real user environment concerning system performance, user perception, acceptance and HRI, guide the development and research during this phase and for the second iterative development and design process that will be validated in the end evaluation. Therefore the KSERA Evaluation Model (described in detail in KSERA deliverable D1.3 "SAR metrics") is focused on the research question "*how will the users accept and use the system?"*.

During the Formative Evaluation there was a strong focus on the

- **Technical system performance** (performance testing),

which is also related to the

- **User acceptance, perception of the social impact and inference to impact on quality of life (QoL)**
- **User experience concerning the human perception of the robot (HPR)**

and which influences those factors. These topics were also analyzed using the metrics described in D1.3. The combination of the different metrics leads to an interpretable result regarding the overall technical and user experience related performance of the KSERA system. This performance is strongly related to conclusions about perceived usefulness (PU) and perceived ease-of-use (PEOU) of the system (see also [Davis1989]).

Figure.1 gives an overview of the different evaluation domains. This approach of combining different domains increases the modularity and hence offers the opportunity for **„flexible testing“** (in terms of possible changes – adaptations - to the system during the trial phases, as single domains can also be used for evaluating sub-testcases (during the end-user trials as well as lab trials)).

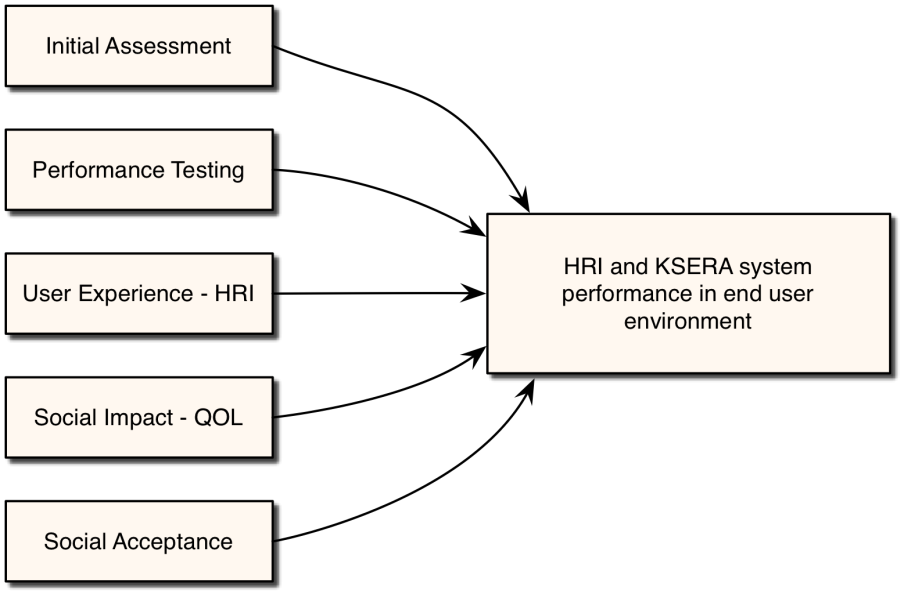


Figure 1: method mix used during the KSERA evaluation phases

1.2.1. System performance

As one of the main factors for the perception of the usefulness (PU) and ease of use (PEOU) [Davis1989] the **system performance** of the integrated prototype 1 system was in the focus of the formative evaluation. Technical functionalities and related performance were evaluated based on their impact on user related parameters concerning:

- Effectiveness,
- Efficiency,
- Flexibility,
- Robustness

using the performance testing metrics described in D1.3 SAR metrics, which were related to several testcases and subtestcases .

1.2.2. User acceptance, Perception of the social impact and inference to impact on quality of life (QoL)

Concerning the evaluation of **QoL** our approach was to identify needs of the later users, which are related to their quality of life in daily life and which are also related to abilities of support and assistance given by the KSERA system. Based on this we identified the three domains

- Medical measurement and ICT based health support abilities
- Exercise and training support abilities
- Disease self management

as the most promising features of the KSERA system for improving the QoL of people with COPD and of older people in general. These domains are also strongly connected to the scenarios described in D1.1 (for the Formative Evaluation: **“Healthy through Exercise”** and **“Disease Self management”**). Based on these

domains measurements concerning the question of the perceived usefulness of the KSERA approach regarding the **acceptance** of the KSERA system and **a possible improvement in the quality of life (QoL)** of the later users were covered with questionnaires (shown in D1.3). During the trial planning process several hypotheses were defined, which were then answered, using the analysis results based on the defined metrics.

Hypothesis concerning the user acceptance and general impression of the system

"…the KSERA approach of a moveable, intelligent user interface offers a number of advantages concerning the human-computer-interaction compared to state-of-the-art user interfaces"
"…Interaction with the used SAR system offers advantages compared to traditional input/output methods"
"…NAO offers better possibilities for provision of dedicated information – also in special situations, for people with special needs"

Hypothesis concerning the abilities for performing physical parameter measurements. By using the KSERA system…

"…there is an improvement concerning the regularity of physical measurements of the users' parameters"
"…there is an improved quantification of parameters compared to the actual status."
"…there is an improved motivation for the use of measurement tools (e.g. blood pressure) compared to the actual status"

Hypothesis concerning the exercises and training support abilities. By using the KSERA system…

"…there is an improved motivation to perform the individual exercises compared to the actual status (only training plan, personal trainer, reminder,….)."
"…there is an improvement concerning the regularity of performing physical training tasks."
"…there is an improvement concerning the training task accomplishment compared to a GUI."
"…the combination of a SAR and video beamer is fruitful concerning the understanding of the spatial movement tasks."

Hypothesis concerning the disease management

"…the KSERA approach can improve advice strategies for a self-disease-management."
"…the KSERA approach can improve emergency strategies compared to stationary user interfaces."
"…the KSERA approach is able to improve the disease awareness."

1.2.3. User experience concerning the human perception of the robot (HPR)

During the formative evaluation as one interesting point concerning the perception of the whole system and the overall human-robot-interaction (HRI) the users' perception of the robot was evaluated using a project internal KSERA scale and the standardized Godspeed scale (also described in D1.3 SAR metrics). Those scales focused on anthropomorphism, animacy, likeability, perceived intelligence and perceived safety

2. Final evaluation test-setting

2.1. Description of Test Users

2.1.1. Trials

The criteria to take part in the formative evaluation have been being at least 70 years old and cognitively healthy. Furthermore the physical abilities should permit to conduct the physical exercises. The users had to agree with the audio and video recording of the trials and signed an informed consent document.

Eight persons with an average age of 77 years (71 – 90) took part in the formative evaluation; seven of them female, one male.

2.2. Test setting

A full featured PT1 installation was undertaken at both test sites in Austria and Israel; the illustration below shows the system components and their location in the test environment within the „Living Lab Schwechat“ in Austria.

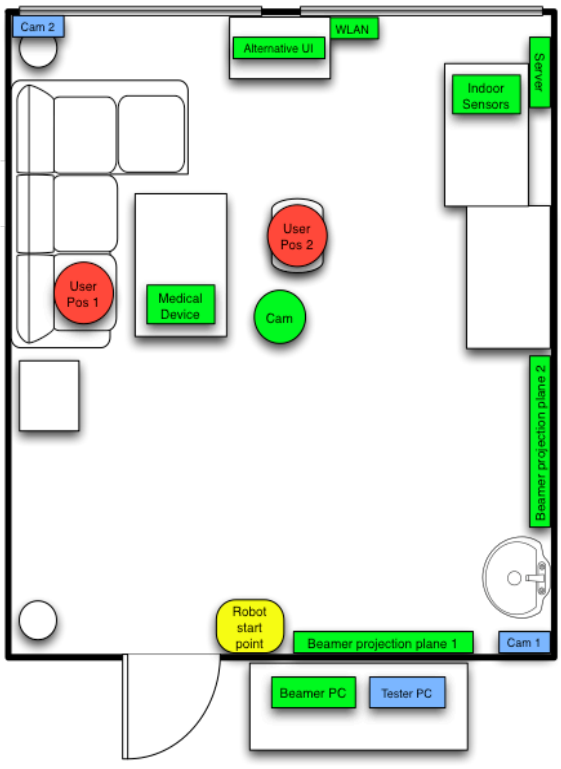


Figure 2: Test room floorplan of the "Living Lab Schwechat”

Figure 2 shows a schematic floor plan of the test setting at the living lab Schwechat. Elements depicted in green are technical equipment needed and used for the trials and are part of the KSERA system.

- Medical device – a pulsoximeter „Nonin 9560“ which was used as an example of a medical measurement device
- Cam – a ceiling cam used for localization of the robot and the user and for navigation of the robot.
- Indoor sensors – indoor environmental sensors for temperature and humidity used in testcase 3

- Server – the KSERA server running the KSERA system
- WLAN – a router hosting the wireless network used to connect the wireless robot
- Beamer PC – a dedicated pc to relay the videos to the wireless video projector mounted on the robot
- Beamer projection plane - spaces that were kept free of furniture to allow the robot to find a free spot on a white wall for projecting videos of exercises during testcase 2

Figure 3 shows the trial site in the Living Lab Schwechat in the senior citizen home Schwechat. The picture gives an impression of the real life condition (concerning furnitur, arrangement, light and sound conditions.)

Figure 3: KSERA Trial setup (Senior citizen center Schwechat)

During the second iteration of the trials in Austria a touch screen based user interface (depicted as Alternative UI) of the FP7-funded project „Long Lasting Memories“ was used for comparison of the KSERA system with a static, non mobile, system covering similar functionalities. This is shown in Figure 4.

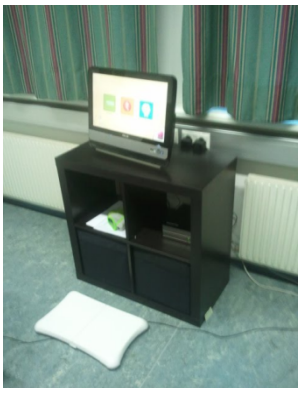


Figure 4: Touchscreen based solution LLM LUI

Elements colored in blue are part of the system used by the researchers to record the testcases and are not part of the KSERA system.

Two user positions are shown that mark the usual positions of the users during the trials. User position 1 was used during the demonstration of testcase 1; user position 2 was used during demonstration of the other testcases.

The robot started all testcases at the yellow marked “robot starting point” aside of the entrance door.

During the test three members of the research team supported the test flow. Two researches (with inside knowledge of the technical system) were located outside the room during the test to steer the startup and stop of the testcases, control the audio and video recordings and to fill out technical performance questionnaires. The third researcher was concerned with the user involvement and stayed with the user in the room in order to be able to support the test user (except during the navigation of TC1 where this researcher stayed in the doorframe to avoid influencing the technical performance of the system’s navigation)

To control the test flow the technical researchers used a technical user interface as shown in the figure below.

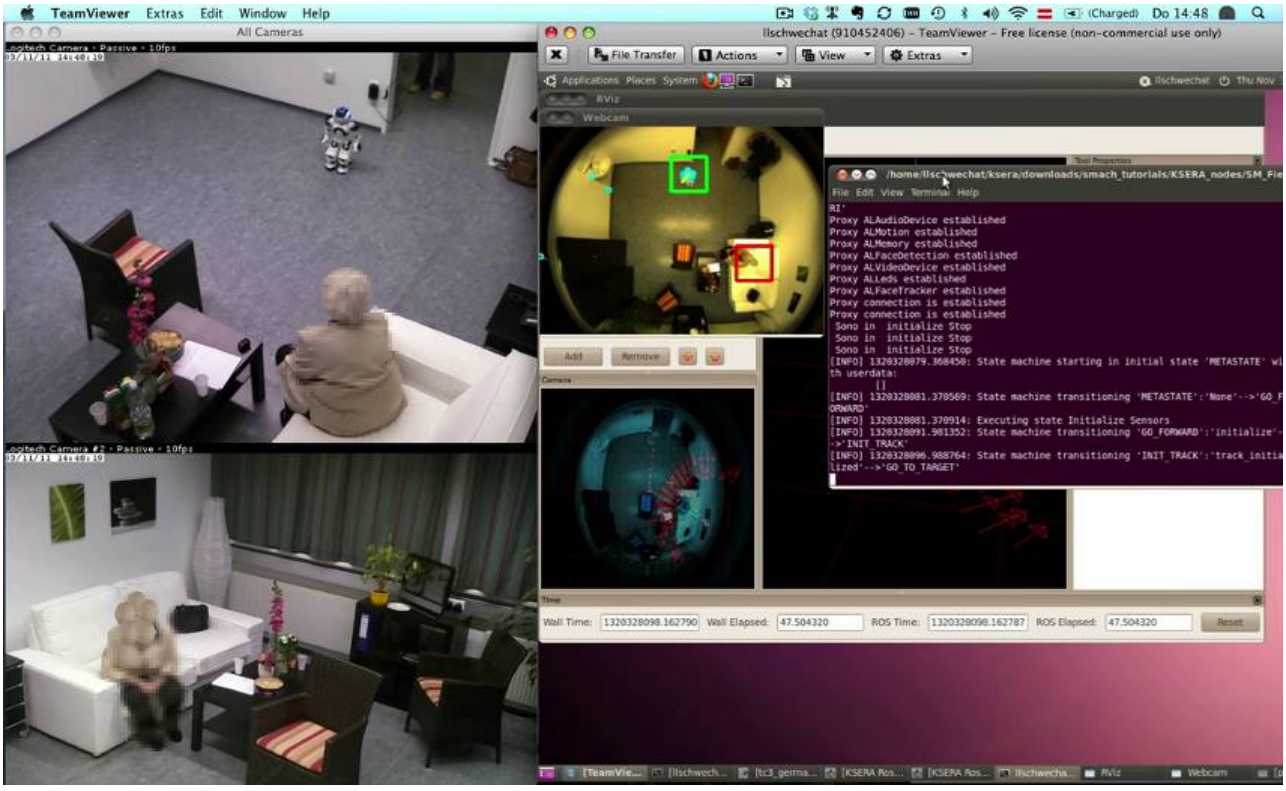


Figure 5: researcher user interface

The researcher interface gives the technical researchers the possibility to observe the scene using 3 cameras mounted on two corners and the ceiling of the room. A connection to the KSERA server enables the direct manipulation of the test flow by using shell commands. This researcher interface was screen captured during all testcases for retrospective analysis. This data and video material was also shared with the other partners in the KSERA consortium in order to realize an iterative improvement to technical problems.

The following figure shows the data flow of the KSERA system and highlights the functionality of the researcher interface.

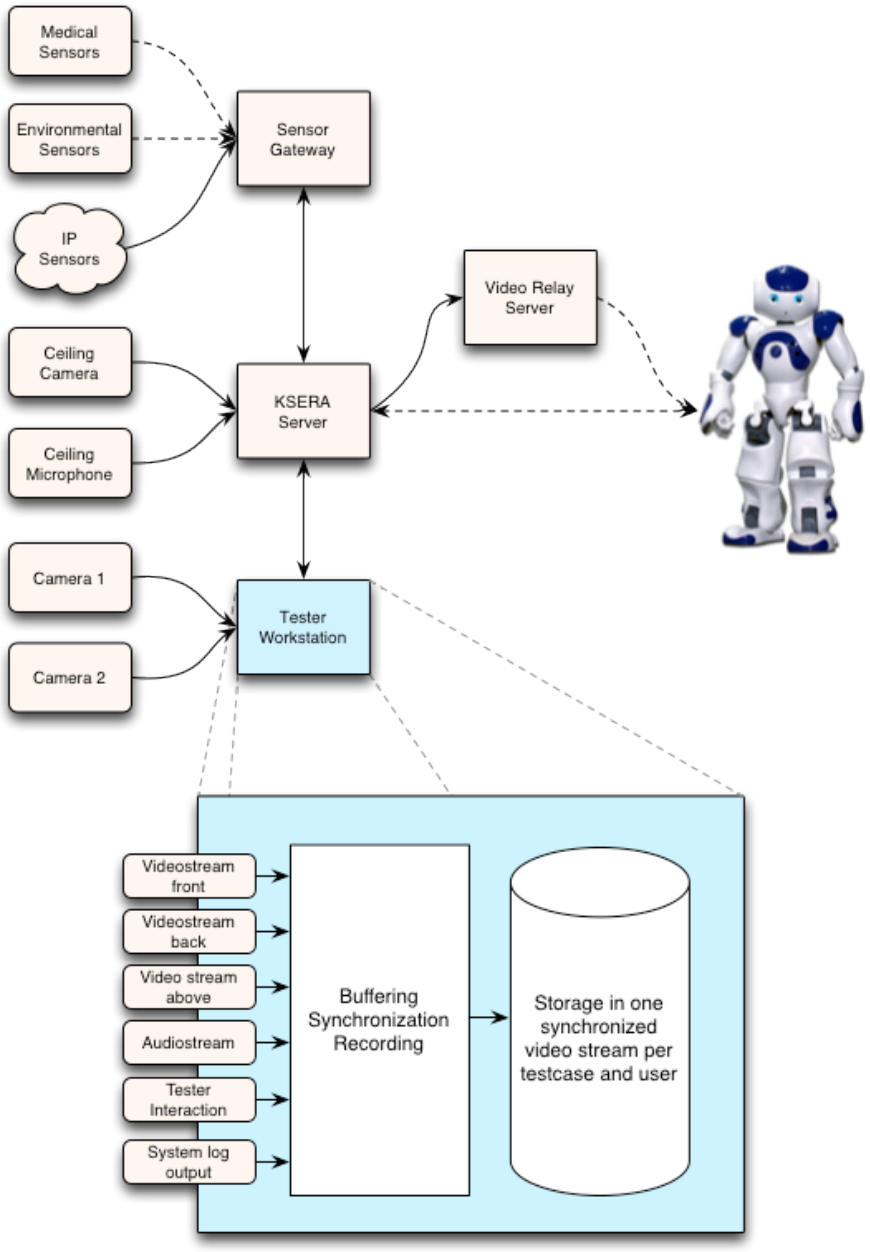


Figure 6: Data flow and synchronization of recordings within the KSERA system

All available video / audio streams, the system logging output and the researcher interaction during the test cases was time synchronized and recorded for later analysis of the trials in terms of human robot interaction and technical performance.

2.3. Test procedure

Each trial was conducted in the following steps (phases). The average procedure for a single test with one trial participant took about 90 minutes on average – including all of the following test phases.

2.3.1. Preparations prior to the test

A checklist was used to make sure that the technical system is ready for the test.

2.3.2. Reception of the user

The project was explained to the participants, his/her role during the tests was discussed and the formal agreement of the participant was given within the informed consent document. This phase was conducted without the robot being visible.

2.3.3. Pre Test Phase - initial questionnaires

To evaluate the current emotional and physiological state of the participant the following questionnaires were filled out together with the user.

The following questionnaires were done in the first iteration:

- WHO – quality of life
- Attitude towards technology
- PANAS

2.3.4. Explanation of the KSERA-system

Every visible part of the system was explained to the user:

- NAO
- Ceiling camera
- Ceiling microphone
- Research cameras / microphones
- Test setup (e.g. user in the test-environment, researcher outside the environment watching)
- Explanation how to handle the pulseoximeter
- Explanation of the optional additional local user interface (only in second iteration of Austrian trials)

2.3.5. Demonstration of test-cases

Three testcases were demonstrated to the user starting from the technically simple ones to the most complex. Each testcase was shown twice in order to gain statistical relevant technical information on the reliability of the system and to give the user the chance to experience different options of the system.

The following testcases were conducted:

1. Medical measurement – the robot navigates to the user, detects the user's face and asks to do a medical measurement using a pulseoximeter
2. Physical training – the robot demonstrates physical exercises in order to motivate the user to do physical training
3. Environmental warning – the robot informs the user about the current environmental conditions using voice output

During the run of the testcases the technical researchers outside of the room filled out the technical reliability questionnaires. Further the „thinking aloud" approach was used and thoughts of the test user were noted.

During the second iteration of the trials conducted in Austria, for each of the KSERA testcases a similar scenario using the alternative touch screen user interface was tested and an interview concerning the preferences of the users between the technical systems was conducted.

2.3.6. Post test phase - final interviews

Directly after the last test run the final interviews took place regarding the following topics:

- User acceptance – general impression
- User acceptance – personal feeling
- User acceptance – satisfaction evaluation
- User acceptance – physical measurement abilities
- User acceptance – disease management
- KSERA - HRI Scale
- HRI Godspeed

2.3.7. Discharge and planning of future tests

After filling out the case report form (CRF) the user was told about the next steps of the project and further participation was planned.

2.4. Limitations concerning the analysis results

Certain technical limitations exist in the prototype 1 (PT1) system that had to be circumvented during the trials to allow meaningful results concerning the gathered user experience data. The following points changed the results of the technical performance evaluation in a positive way and hence limit the significance of certain parts of the technical evaluation.

- Light conditions were set to be as static as possible. Artificial lights were used, external lights were omitted by closing the curtains and using window blends. → This Influences the localization, navigation and face detection results positively
- During beamer projection the artificial lights were turned off → influences the quality of the projected image positively
- The initial positions of the robot, the user, the final projection spot (target spot) where fixed prior to the trials to achieve comparable results between iterations → influences the localization (user and robot) and navigation positively
- The user was kept active during the navigational part either by giving small tasks, such as „please can you take the pulsoximeter“ or „take a look in the newspaper“ in order to allow the technical system to locate the user by the user's movements. This was necessary to conduct the trial but influenced the technical performance of the navigation component in a positive way when compared to a real-life situation → influences the user localization, navigation positively
- User and robot positions where carefully chosen to avoid obstacles and allow for a simple path from the robot's starting position to the desired end location → influences the navigation positively
- Keyinput was used instead of word recognition → this did NOT influence the recognition rate of the word detection since this measurement was done before the key input and the value was taken from the system logs

2.5. Changes during the test phase

During the test the system was kept static to ensure comparable results between iterations. After the first two tests in Schwechat the following minor bugfixes were implemented in the systems in both test sites:

- The german pronaunciation of decimal numbers (only in Schwechat)
- The starting position of the robot was changed minimally to get better results in TC2 (the angle of the starting position was slightly adapted)

3. Results – evaluation report

3.1. Test-statistics

Prior to the Formative Evaluation trials pretests in the test sites were performed, which can be compared to lab tests in order to guarantee a working system and to evaluate the trial procedure and test workflows. A pre-test with an older user was performed in order to check limitations concerning sound output, input, visual recognition, etc. This procedure helps to gather comparable results from the first to the last trial participant.

In the trial phase together with older users (age 70 to 91) a group of 16 trial participants (8 in Israel, 8 in Austria) joined the tests. They were invited for an initial conception of the tests and the related procedure (but without showing them the system or the test site).

The test plan included two iterations at every test site. As example of the gathered data the test statistic of the tests performed in the living lab Schwechat is presented:

- 8 users / two iterations
- 3 testcases per iteration + comparison with actual smart home solutions
- 8*3*2+8*1= 56 single running testcases = videos of 3 cameras (ceiling, front, back) and research interface
- Over 600 single evaluated technical subtestcases (metrics)
- Over 500 minutes of human robot interaction and research video material
- Questionnaires and qualitative information sheets

3.2. Results of the Technical Performance Evaluation

This section describes the results of the technical evaluation in terms of

- robustness,
- efficiency and
- effectiveness

of the first KSERA prototype. The results were gathered using predefined performance metrics. The researchers used predefined performance criteria and thresholds to evaluate all of the metrics. Most metrics (except the duration measurements used to evaluate the efficiency of the system) used a binary true / false decision as outcome that could be evaluated by the researchers during the running tests in time to ensure high accuracy by gathering the data live in comparison to a later video analysis.

3.2.1. Localization, Navigation and Movement

Performance of the localization metrics

The KSERA system uses localization for two mostly independent localization tasks: Localizing the user and localizing the robot within the test environment. For both tasks a visual approach was applied facilitating a ceiling mounted camera. Particle filters were used to detect moving targets and pattern recognition to differentiate between the user and the robot.

The following metrics were used to evaluate the localization performance:

- A3 - „Is the KSERA system calculating the right user position“
- A4 - „Is the KSERA system calculating the right NAO position.

Both questions were evaluated during the test by interpreting the research interface as shown in the image below.

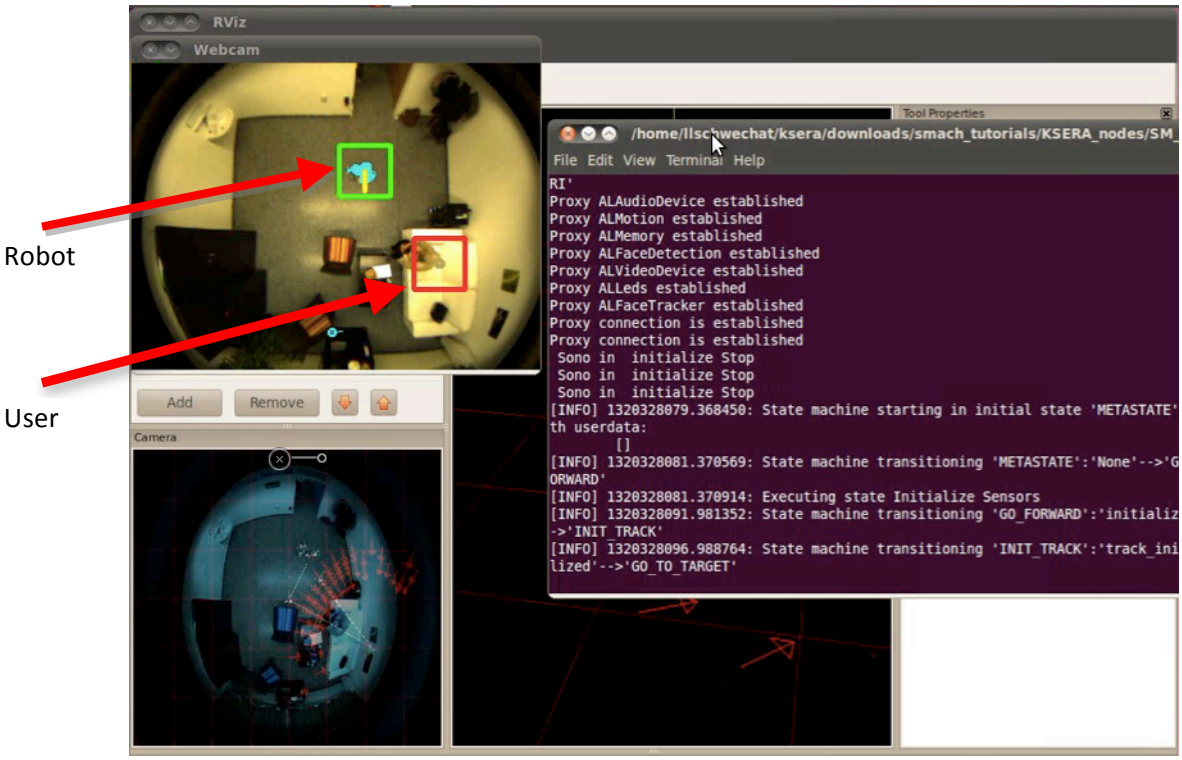


Figure 7: technical research interface for evaluation of the localization performance

In the upper left part of this screenshot a head view gathered by the ceiling camera is depicted. The green and red squares in this view show the estimated location of the user (red) and the robot (green). The yellow line inside the green square indicates the calculated direction of the robots head. Defined criteria were used to evaluate whether the system met the requirements, in this case the markers size is equivalent to approximately 1m by 1m; this subtestcase failed if the marker did not cover the trial participants or the robot respectively at all, indicating a deviation by more than 0.5 meters.

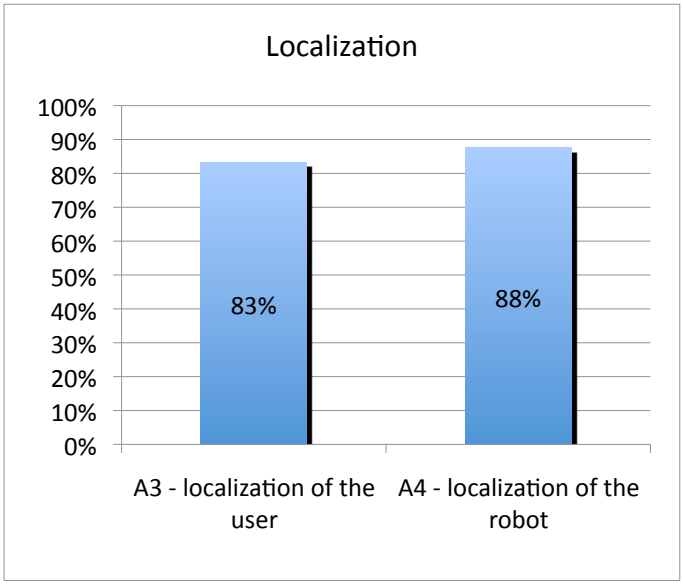


Figure 8: summary of localization metrics results

The figure above shows that in 83% of conducted subtestcases (n=24) the system correctly located the user. The localization of the robot correctly worked in 88% of subtestcases (n=32).

Since the performance of the user localization varies with the degree of movement of the user, the user was asked to walk into the room at the beginning of the system's localization. In most erroneous cases the system initially found the user during this walk, but then lost the position again after the user stayed rather stationary on the sofa. The robot localization had a particular problem in misinterpreting the direction of the robot by 180°.

Performance of robot movement metrics

During the testcases robot movement was used in different scenarios and four different subtestcases:

- to navigate the robot to the user or to a specific spot in the test environment
- to turn the head to find the user with the head mounted camera
- to demonstrate training exercises
- to raise the user's attention by waving with the arms.

Since different movements were used for these respective subtasks, the evaluation metrics were split into subparts for each motion.

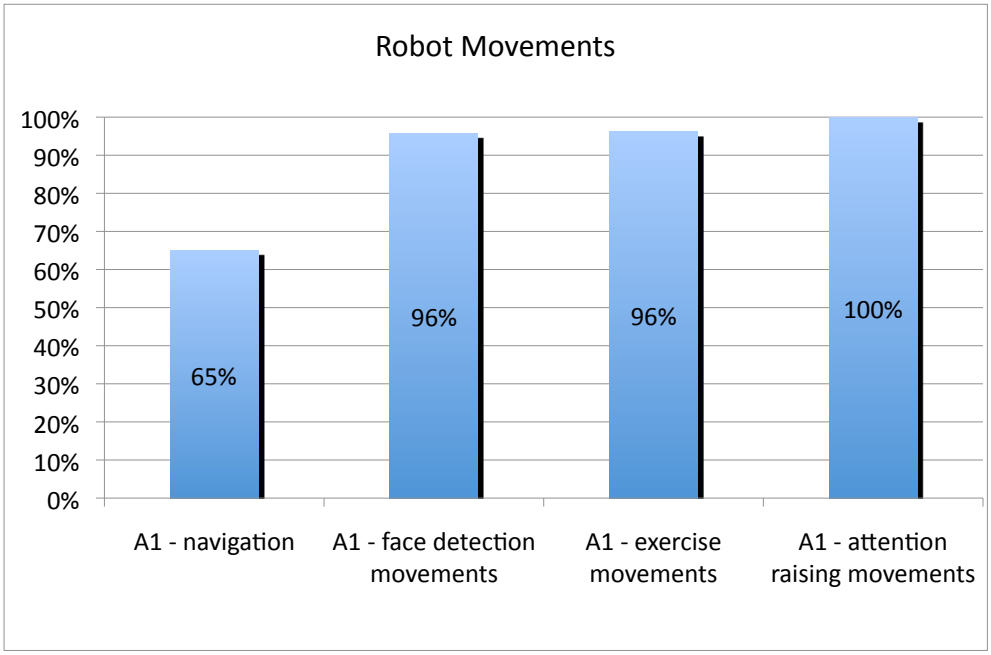


Figure 9: results of robot movement evaluation

Each metric was evaluated by an optical live analysis and interpretation of the robot movement during the testcase and, if needed, by using the recorded video material retrospectively.

The movements of the robot itself worked very reliably however were negatively influenced in the case of navigation by the localization subtasks of the system.

In 65% of the cases (n=38) the robot navigated correctly to the desired spot, in the rest of the cases mostly the localization of either the user or the robot did not work in a way that let the robot reach the target location as intended. In testcase 2 the most common reason for a failure was the incorrect alignment of the robot to the wall to allow projecting an image to the wall. In testcase 1 the most common problem was the localization of the user, which lead to a wrong destination point. The preprogrammed walking movements were correct in nearly all cases, except for one time when the robot lost balance and crashed on the floor during a walk. The reason for this problem is unknown but possibly the additional weight of the, at his time, back mounted beamer played a role.

The face detection movements (n=24) , attention raising movements (n=23) and exercise movements (n=27) scored with high results which shows that the system is capable of performing even complex movements as used during the demonstration of physical exercises reliably.

Figure 10: physical exercises demonstrated by movements of the robot

The figure above shows the robot guiding and motivating a user to do physical training by performing exercise movements.

3.2.2. Performance of Interaction and communication metrics

Performance of voice communication

The SAR uses voice communication to inform the user and the user is able to command the robot by voice input as well. In this first prototype only the phrases “yes / yes, please” and “no / no, thanks” were implemented to be used by the test users to tell the system whether or not they complied with the task the robot gave them.

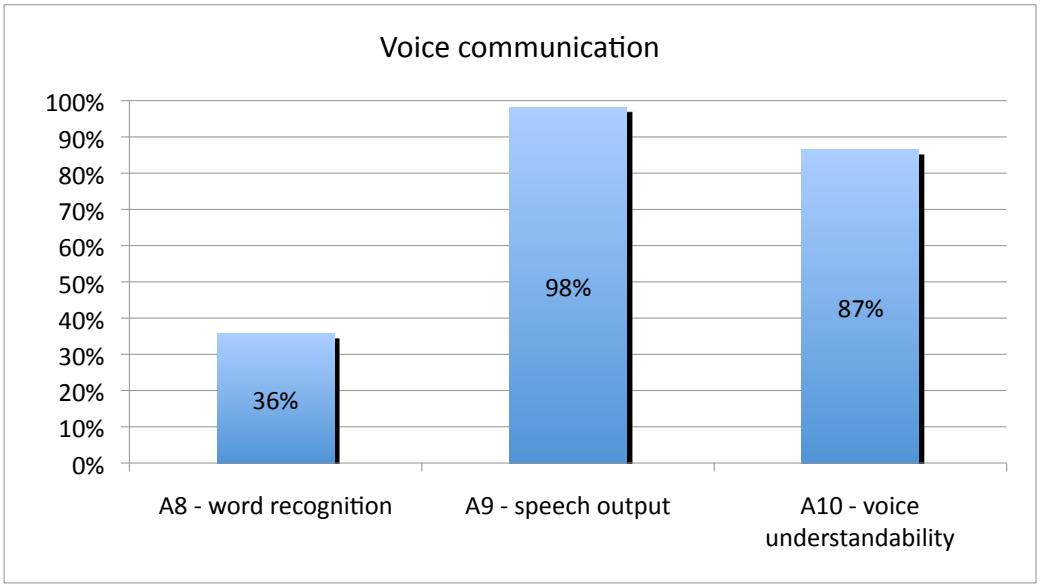


Figure 11: performance results of voice communication

The respective Figure 11 shows low performance results of 36% (n= 28) of the word recognition that were already known prior to the field trials after system testing but could not be solved in time for prototype 1.

Because of this limitation the researchers used key input to simulate the word recognition to avoid influencing the user experience results negatively. The low performance is partly due to the fact that the microphones at the test sites were not placed on the user directly but about 1m away from the user. The bad performance was due to the very complex sound situation in real life environments and the request by the trial sites not to use body fixed microphones, which would have the positive effect of a better sound perception for the system, but also would be a solution that has to be worn on the body and thus introduces usability issues concerning a planned future use at home.

The speech output metrics nearly always performed successfully, scoring in 98% at a total of 212 evaluated cases (n=212), in few cases at the beginning of the trials, the German version had some pronunciation issues that could be solved during the tests.

The voice understandability was measured by interpreting the reaction of the user on voice output of the robot. In many cases the users also either responded correctly to the SAR or directly claimed that the could not understand the robot. The results show a performance of (87% at n=216), which hence was not optimal for some users and varied a lot between the test-users. Interestingly, the understandability could not be improved solely by turning up the volume of the system to its maximum. It seems this parameter is partly influenced by other factors such as intonation and speech melody of the computer generated voice as well as vocabulary used.

Performance of further HRI metrics

The following figure shows four independent metrics grouped into one graph.

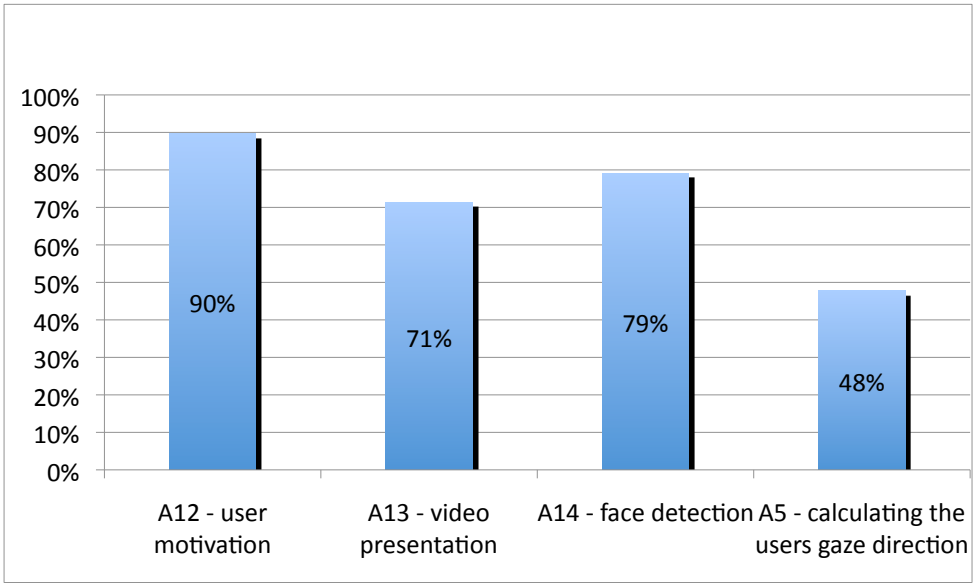


Figure 12: performance evaluation of four HRI metrics

The metric A12 – user motivation refers to the ability of the KSERA system to motivate the user to perform the medical measurement, this was assessed as being correct in case the user did perform the measurement when asked to do so by the system. In 90% (n=39) of the cases the users did take the pulsoximeter and performed the task. In the rest of cases the user did not follow the robots commands; mostly because the user was either distracted or did not understand the robot. None said that he/she is not willing to do so.

The metric A13 was used to determine whether or not the video display functionality of the SAR was working correctly. This functionality was used in testcase 2 to show exercises to the user by projecting prerecorded video material with the mobile LED projector mounted on the robot onto a clear spot on the wall. See also the figure below.

Figure 13: robot displays physical exercises with an attached video projector

The quality of the projected image was largely dependent on the light conditions in the room and the distance between robot and the wall and was not included as factor in this metric. This metric only covers the question whether or not the correct video content could be projected to a wall. In 71% of the cases (at only n=7) this was evaluated as true. In two of seven cases the test had to be aborted because of a freeze of the program used to relay the video to the beamer on the robots back.

During the trials also a split into subtestcases was performed, where the overall testcase of moving NAO through the environment, finding a wall for projection, moving itself into a distinct beamer position (so that the beamer axis were exactly in parallel with the wall) and then projection of the picture, where viewed and performed in single test sessions, which offered the opportunity to better evaluate the projection performance.

The metric A14 was used to evaluate the face detection capabilities of the system. In case the system could not find the user's face during the face searching routine or mistakenly identified a part of the room as face, this metric failed. In 79% of cases (n=24) this metric could be evaluated positively. In most other cases the users wore glasses that reproducibly led to a failure of this routine. In two cases the system mistakenly detected a face in the curtains and in the drawer respectively.

The system estimates the gaze direction during the voice communication parts of human robot interaction in testcase 1 to determine whether or not the user is paying attention to the robot. In 48% of the cases (n=23) this algorithm seemed to misinterpret the head orientation leading to additional movements for attention raising. This metric was treated as not fulfilled in case the robot undertook attention-raising movements although the user was clearly focused on the robot. In 48% of cases at a total of n=23 this worked fine.

3.3. Performance of data gathering metrics

The KSERA system uses external interfaces such as Bluetooth and HTTP to gather environmental and medical data.

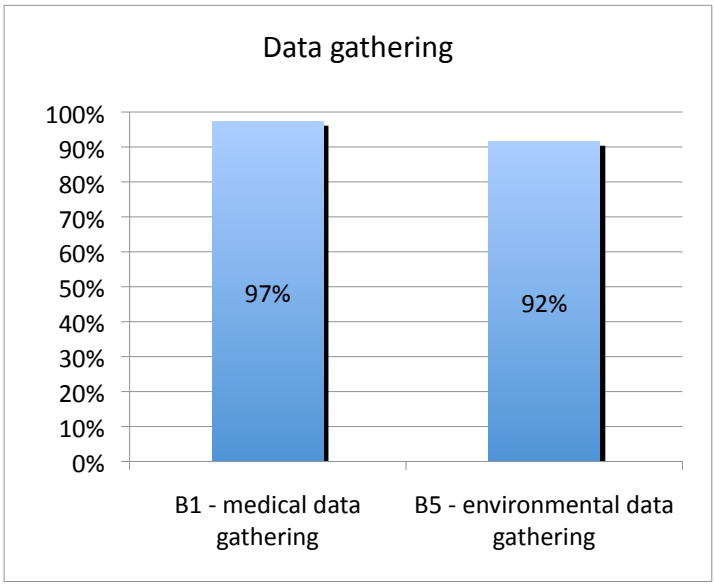


Figure 14: performance results of data gathering

It could be shown that data gathering from the used pulsoximeter and the used indoor and outdoor sensors worked mostly reliably during all tests at the testsites. In rare cases either the connection to environmental sensors which relies on a proper internet connection, bandwidth and latency or the Bluetooth connection to the medical measurement device was broken resulting in either no or incorrect reported values.

3.3.1. Evaluation of the duration of tasks

To evaluate the efficiency of the technical system in delivering content to the user, time measurements of task durations were undertaken as part of the metrics. The following figures show the time taken by the system for tasks that are not directly used for human robot interaction. The time is given in seconds.

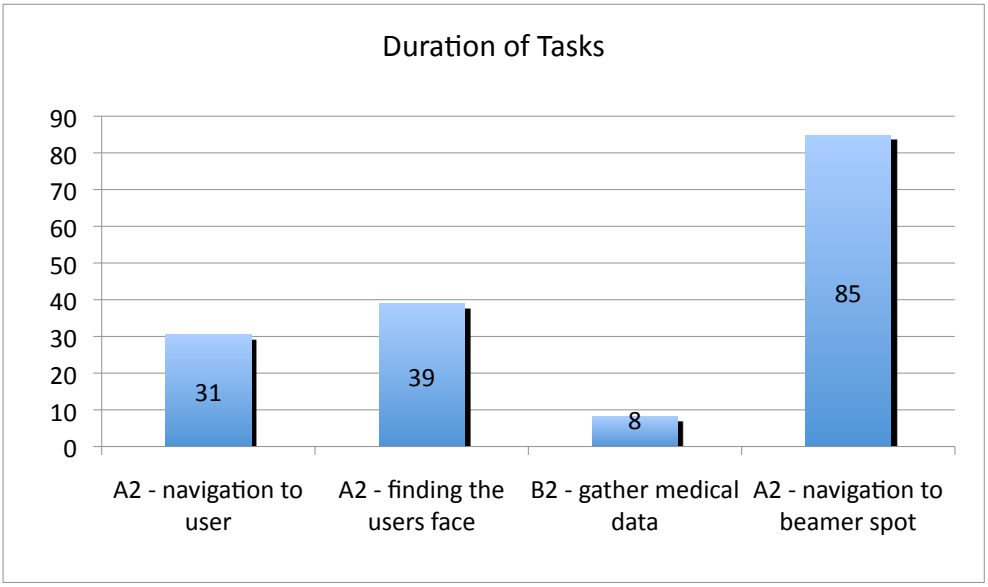


Figure 15: performance results on the duration mean in seconds of certain tasks

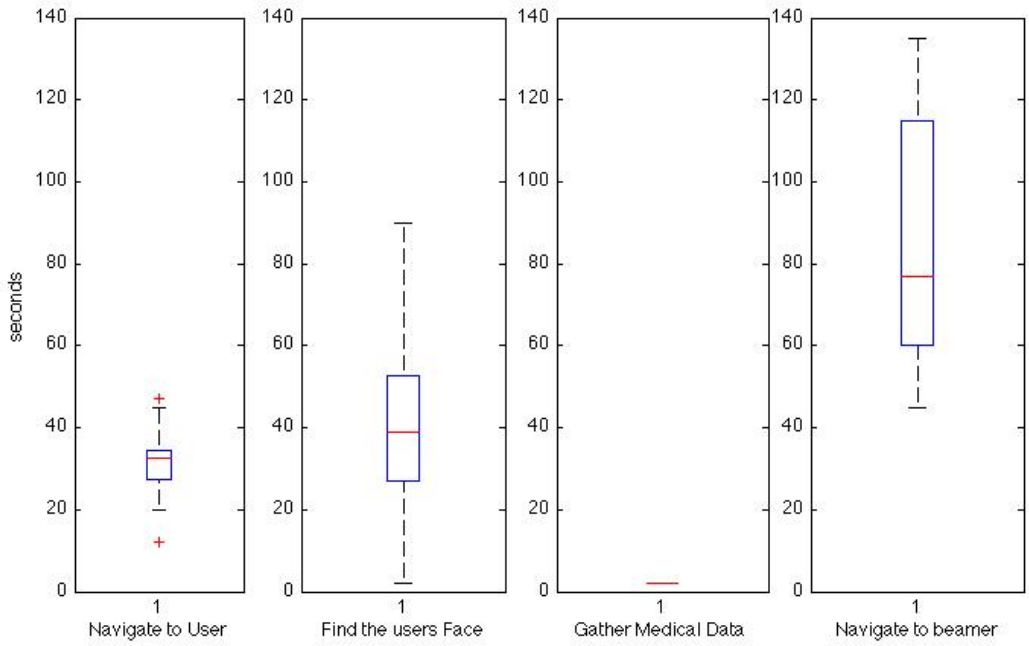


Figure 16: duration analysis of tasks

The tasks „navigation to user“ and „finding the user’s face“ are performed in the beginning of testcase 1 to walk to the user and initiate voice communication. Together these tasks take in average 70 seconds (n=38, n=19), which is the time it takes the system to start communicating to the user. This rather long time is partly the reason why some of the users reported the system as being „slow“.

As shown in the boxplot diagram in Figure 16 the time measurements vary a lot between different iterations. The reason clearly lies in the low repeatability of the navigation process. The following figure shows the trajectories gathered by retrospective analysis of the gathered video material of two extreme cases.

Figure 17: comparison of navigation performance between iterations (NAO walking to the beamer position)

The difference in performance between taking the direct way or not was not reflected in the navigational metrics but in the duration taken until the robot reached the target position.

Gathering medical data was evaluated using only one medical device, the Bluetooth pulsoximeter. This device very reliably gathered data quickly and sent it to the KSERA system without a mentionable delay, the researchers measured an average of 8 seconds at n=34, this is hence working well.

The metric A2 – "navigation to beamer spot" took the most time and hence negatively influenced the overall system's efficiency. It took nearly one and a half minutes for the robot to reach the designated spot for beaming (n=6) which was located about 2 meters away from the robot's starting position without obstacles in between.

3.4. Summary of the technical statistical evaluation

Some metrics could not achieve the expected rating due to real life circumstances, however within the real-life simulated environment most technical issues could be concealed from the user to allow valid test results from usability testing.

Systematic errors (such as attracting the user's attention at the wrong time) and errors driven by external conditions such as light, motion in the room (navigation) and external sound sources (speech recognition) could be found and reported to the technical team for improvement for the second prototype of the KSERA system.

The figure below shows a summary of the evaluation statistic clustered by the three conducted testcases. The given numbers state the likelihood of an average subtestcase to function correctly within this testcase.

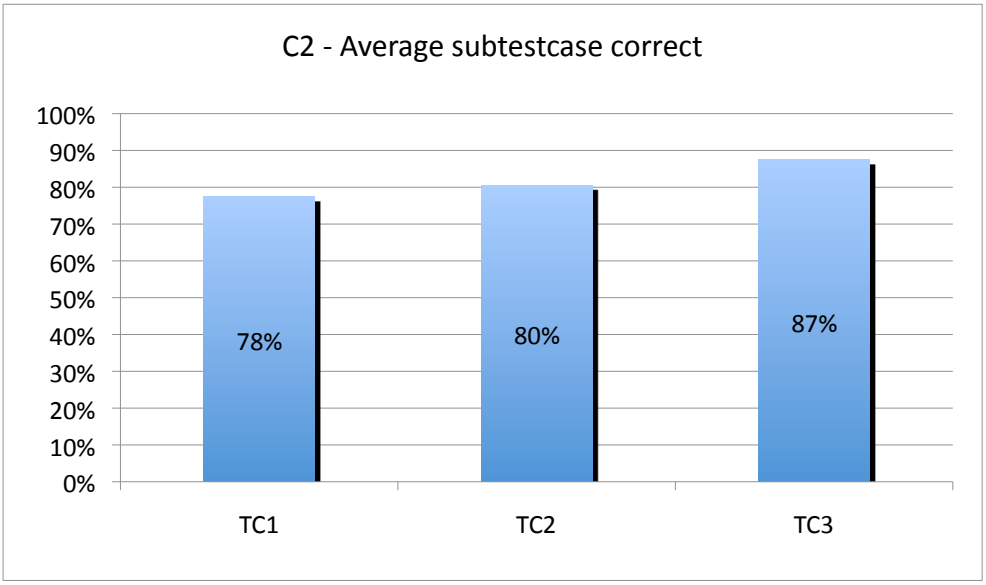


Figure 18: performance summary clustered by testcase

As can be seen testcase 3 has a higher reliability as the other test cases. This is mostly due to the fact that no navigation is needed which would involve highly complex technical procedures such as pattern recognition and object tracking.

The following table provides the full evaluation statistics that are described in this chapter as reference.

	TC1			TC2			TC3				
Metric	Sum	n	%	Sum	n	%	Sum	n	%	Avg	n
A1 - navigation	31	47	66%	10	16	63%				**65%**	63
A1 - face detection movements	23	24	96%							**96%**	24
A1 - exercise movements				26	27	96%				**96%**	27
A1 - attention raising movements	23	23	100%							**100%**	23
A2 - navigation to user	31	38								**31**	38
A2 - finding the users face	39	19								**39**	19
B2 - gather medical data	8	34								**8**	34
A2 - navigation to beamer spot				85	6					**85**	6
A3 - localization of the user	20	24	83%							**83%**	24
A4 - localization of the robot	21	24	88%	7	8	88%				**88%**	32
A8 - word recognition	0	8	0%	10	20	50%				**36%**	28
A9 - speech output	95	97	98%	70	70	100%	43	45	96%	**98%**	212
A10 - voice understandability	85	99	85%	68	71	96%	35	46	75%	**87%**	216
A12 - user motivation	35	39	90%							**90%**	39
A13 - video presentation				5	7	71%				**71%**	7
A14 - face detection	19	24	79%							**79%**	24
A5 - calculating the users gaze direction	11	23	48%							**48%**	23
B1 - medical data gathering	37	38	97%							**97%**	38
B5 - environmental data gathering							22	24	92%	**92%**	24
C1 - Testcase entirely correct	7	44	16%	22	38	58%	37	46	80%	**52%**	128
C2 - Average subtestcase correct			78%			80%			87%	**82%**	901

Figure 19 full performance statistic data

3.5. KSERA QoL Scale - Usability Evaluation

As described in chapter 1.2.2 our approach in evaluating the (future) abilities of the KSERA system concerning the enhancement of the quality of life of older people in general and persons with COPD in special was based on giving answers to the defined hypothesis and the related metrics. Information was gathered using the KSERA QoL questionnaire, which dealt with main topics of the

- general impression of the KSERA system and the overall approach
- emotional feeling and satisfaction concerning the system and its functionalities
- evaluation of the meaningfulness of the disease management abilities
- evaluation of the meaningfulness of the SAR robotic training support

Beside these quantitative parameters, also qualitative information was gathered deploying a "Think Aloud" approach. In the following the different main topics will be discussed and will then be connected to the above hypothesis. Furthermore the results from the questionnaires will be connected to the system performance (from the technical point of view) in order to conclude towards possible enhancement in the system and system design and point out the lessons learned.

3.5.1. General impression of the system

The overall impression of the system was identified as very good. On the one hand all participants were quite impressed of the abilities of the system and expressed their feeling, that the system seems to be quite easy to use and to be quite understandable also for beginners and thus easy to learn (as seen in Figure 20: Boxplots of questionnaire data (n=32) and included cluster values of the overall impression of the KSERA approach and the system (maximum value = 5)). The supporting abilities in daily life were seen in a more critical way, however also reached a median value of 4 and thus can be described as above average. The topic of privacy and related topic of having concern about feeling monitored by a mobile agent in the own flat was identified as a minor issue for the test participants (value of 1 indicates "very concerned", 5 "not concerned").

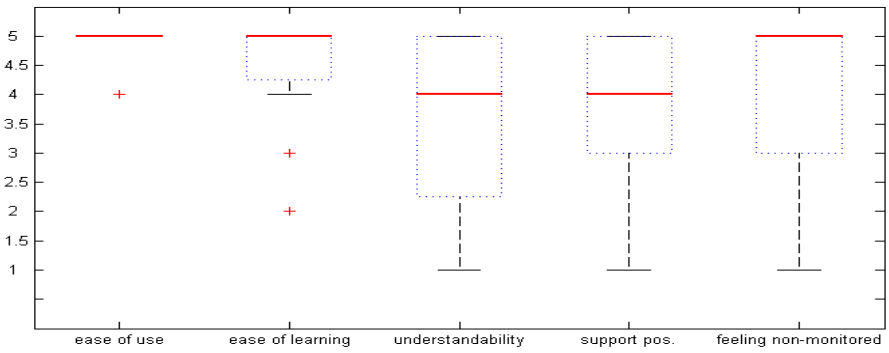


Figure 20: Boxplots of questionnaire data (n=32) and included cluster values of the overall impression of the KSERA approach and the system (maximum value = 5)

This general impression of the system and the related parameters approve the approach to use a humanoid (SAR) robot and the connected ambient sensors of KSERA in an integrated system for daily support in the older peoples' homes and are a base for following in-depth analysis of different system functionalities and the users' perception concerning those. Also the qualitative information was quite positive after the first iteration (an thus first contact with the KSERA system). A summary of positive and negative statements can be found in the appendix "qualitative user comments"

3.5.2. Emotional feeling concerning the system usage, test flow and presented scenarios

This section describes the results of the analysis of the subjective emotional feeling of the trial participants concerning the KSERA system and its performance. These factors are strongly related to parameters describing the possibility for a long term use and long term motivation caused by entertaining systems (*Brickmore2005*, Bartneck2003).

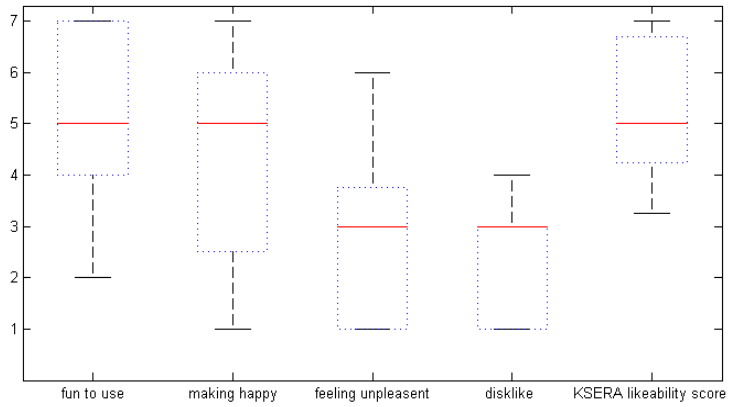


Figure 21: Cluster values of the emotional feeling concerning the KSERA approach and the system (maximum value = 7, n=32)

As shown in Figure 21 the overall perception of the system and scenarios was described as making the people feeling happy – they had fun with the system and only had a minor feeling of unpleasantness or dislike concerning different functionalities. An overall likeability score (giving information about how the person feels about KSERA - corresponding to how the person likes KSERA with a median of 5 with a maximum of 7 was achieved, which can be interpreted as a very positive value for this parameter and thus the system was perceived as **likeable**.

The system usage and the showed scenarios caused no stress or made the participants tired or bored. In contrast the trial participants expressed to be feeling strong, fresh and calm with very high ratings. The score for the feeling concerning the scenarios and system usage after the test resulted in a median of 5.9, which can also be interpreted as very high concerning the overall perception of the system.

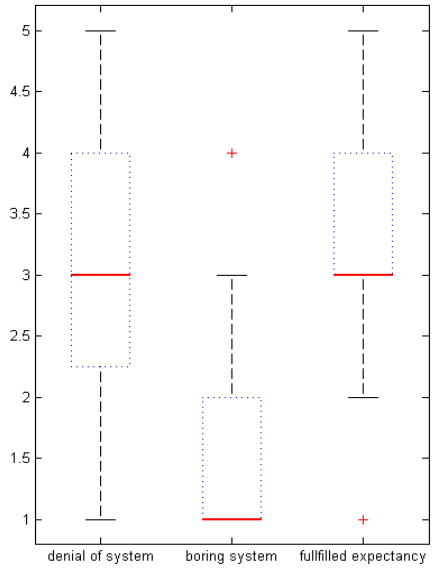


Figure 22: Cluster values for the emotional feeling concerning the KSERA approach and the system (maximum value = 7, n=32)

These parameters are strongly related to the test and trial accomplishment and therefore the conclusion concerning the feeling towards the system must be connected to that fact. Especially the factor of fulfilled expectancy is also related on the one hand to the overall picture of robots (conveyed by the media, films, books, etc.), on the other hand by the information given to the trial participants by the researchers before the trials. The factor concerning the boring system is also influenced by the relatively short human robot interaction sequences and therefore has to be distinguished to same factor results in related studies (e.g. Koay2007), due to the lack of long term interaction.

3.5.3. Level of satisfaction concerning the system functionalities and performance

Figure 23 shows the different categories of (subjective) satisfaction concerning the system functionalities and performance. In the following the different parameters and dependencies are discussed.

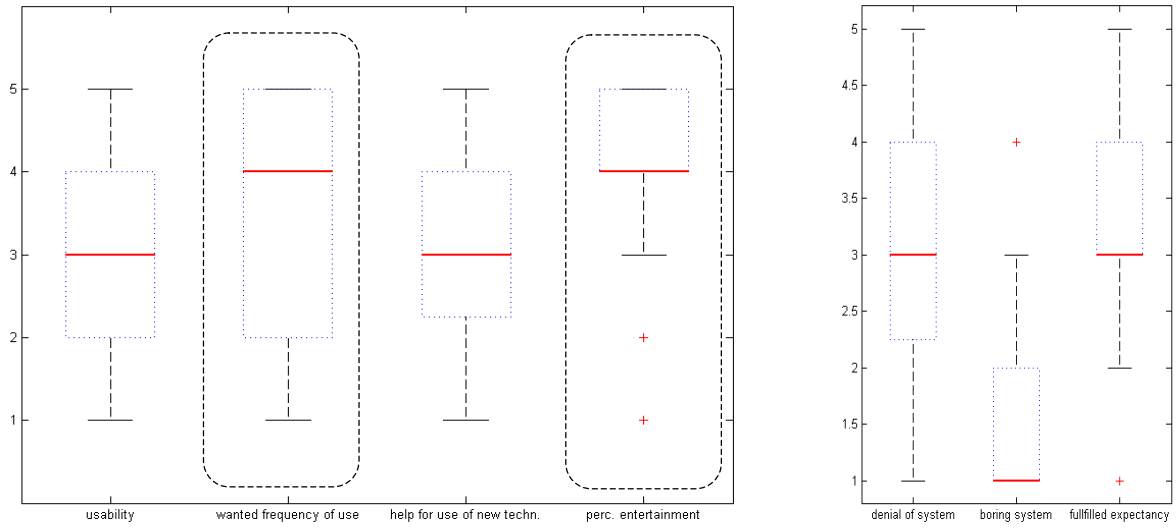


Figure 23: Boxplot of the cluster values of the satisfaction concerning the KSERA approach and the system (maximum value = 5, n=32)

Although most of the trial participants do not think they will need the system in the near future at home or would use it instead of doing other daily activities, they rated the frequency of use, if they would have the system at home very high (median value = 4) – means they would use the whole system or special functionalities (qualitative information about the functionalities can be found in the ANNEX - “Qualitative user comments”) almost every day.

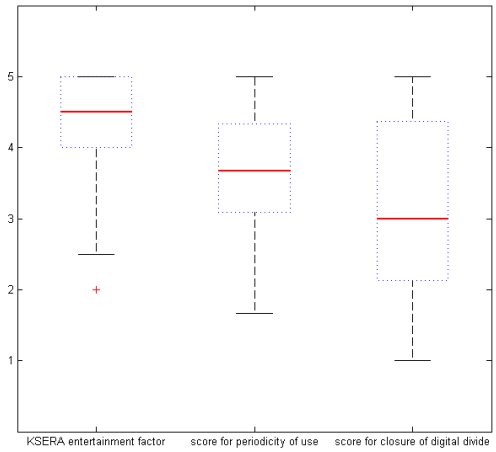


Figure 24: Boxplot of the cluster values of the entertainment, periodicity of use of the KSERA system (maximum value = 5, n=32)

Also the interaction and entertainment abilities shown by the system were rated very high (after the first and second iteration) and it was also described as not boring. Also this entertainment factor correlates with the result of the emotional feeling about the system and will most probably turn out to be a significant parameter for a future regular use of the system, as also seen in many other domains like exergaming (Castellano2009).

An important parameter concerning the information and information preparation for the trials is the "expectancy parameter", which is related to the adequate information process towards the trial participants and was also rated positive. However many of the participants stated, that they expected more humanoid behavior e.g. in the communication with the system, which would be a great benefit – the system and the SAR in special was not seen as a normal computer system – it seemed to be more than just a technical system like a PC (see also qualitative quotes in the ANNEX - "Qualitative user comments").

3.5.4. Physical measurement performance

During the formative evaluation the question for a possible use of the KSERA approach for a future self-disease management was evaluated as well. The results showed a distinct need for self disease-management systems and also very positive feedback concerning the KSERA approach, which seemed to be very beneficial for the trial participants. Figure 23 shows the results concerning the user needs for technical support for disease management in daily life and concerning the KSERA system to motivate them to use such a system.

The results also show that the KSERA approach (combining a mobile SAR with physical parameter measurement sensors and environmental sensors) seemed to be very realistic and promising for a future use.

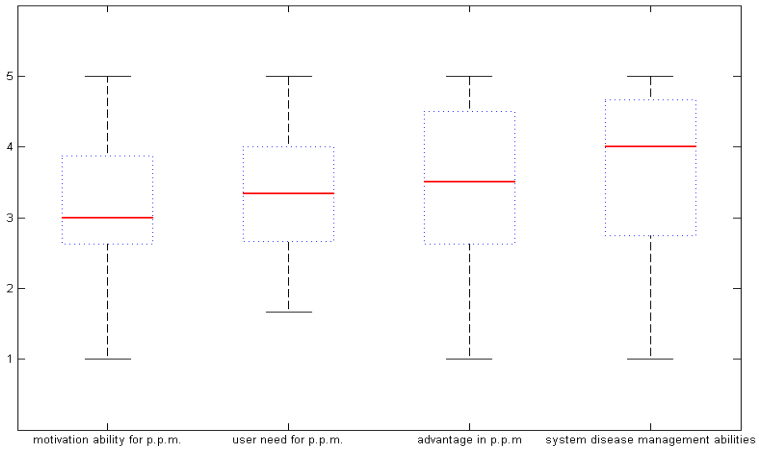


Figure 25: Score values of the physical parameter measurement performance concerning the KSERA approach and the system (maximum value = 5)

This combination of the user need for monitoring support and the motivational abilities of the KSERA system represent a clear hint for ongoing development of this strategy for a future home use of SAR supported disease self management systems. Especially the "humanoid" factor was pointed out several times. It seemed to be very helpful for the trial participants to have a personal motivator at home, who reminds them to perform physical measurements and also deal with the physical measurement data. Many of them stated, that it would be very helpful to have such a system instead of writing down their results and most probably lose them over time. This combination of the technical support for a usual daily performed activity (measurement of physical parameters) and the very personal reminding and support by the SAR was emphasized several times.

During the test session especially the presentation of data was discussed together with the trial participants, as the KSERA system offers different opportunities for representing the results e.g. from physical parameter measurements. The oral presentation by NAO was appreciated e.g. for a direct feedback right after the measurement as it seemed to be a "normal" way of representation like a carer would do the measurement. The trial participants also trusted the system and its abilities of interpreting the data. However also the wish for long term presentation of data was stated very often and also wished for the presentation. In this context a more visual way was preferred (e.g. for showing a trend – for example with a diagram presentation). In conclusion the KSERA approach concerning disease management was appreciated by the trial participants, however additional features would be appreciated. A more complex way of representation (not only the robot speech output) could give an enhanced feedback about the user's status. In this context NAO could be a tool for

- direct and fast feedback,
- motivating the user to perform e.g. physical parameter measurement tasks regularly
- and for instructing the user e.g. how to interpret for example visually presented data, leading the user in the interpretation process.

In short a multimodal way of representation (SAR, screen, LED projector, etc.) was preferred over a standalone – mono-system solution.

3.5.5. Training support

Figure 26 and Figure 27 show the results from the evaluation of the training support abilities of KSERA and especially the supported physical training with the SAR. The results concerning the motivation for task oriented training (TOT) are very promising. Almost all test participants gave a very positive feedback concerning their motivation of performing exercises together with the robot and were very willing to imitate the exercises shown by the robot. Concerning the comparison of actual methods of training support and training explanation, the results were divided. The comparison with "not-human-supported" methods (e.g. written trainings plan) showed, that the KSERA approach was preferred and received rave reviews. The comparison of the robot showing (through physical movements) different exercises and the robot carrying a beamer which displays a video showing the same exercises, showed, that there was a clear preference for the robot showing the exercises by physical movements.

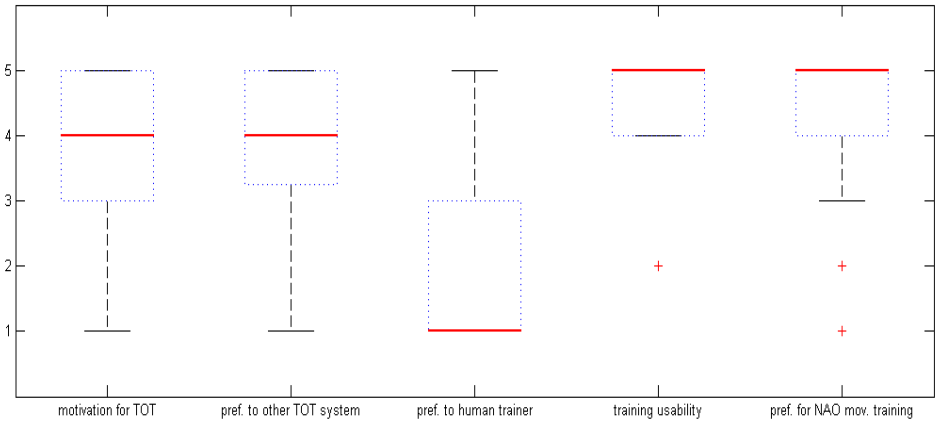


Figure 26: Cluster values of the training support performance concerning the KSERA approach and the system (maximum value = 5, n=32)

However this result could also be influenced by the more complex interaction of different technical modules (which all have to work together perfectly: e.g. NAO navigation, NAO movements, recognition of wall and wall distance, beamer distance, etc.) during the performance of the described testcase. During the trials each test case was split into subtestcases. The overall testcase of moving NAO through the environment, finding a wall for projection, moving itself into a distinct beamer position (so that the beamer axis were exactly in parallel with the wall) and then beaming the picture has been viewed and performed in single test sessions.

This approach showed that the (technical) performance of the whole testcase and the included modules were of a very high importance. When NAO was positioned perfectly and the video picture therefore was beamed in the best way (with no distortion due to bad positioning) the trial participants appreciated having also a visual (in terms of a video picture) presentation of movements – especially for showing the more correct way of movements (due to the limitations in NAO's degrees of freedom).

As an expected result the trial participants preferred a human trainer to the robot, but also stated, that the robot would be a very helpful motivator at home for everyday training, due to its high usability and very clear explanation of movements (in terms of which movement should be performed).

They also said, that it would be good to have human training sessions, where exact exercises would be explained and showed and then be repeated at home together with the KSERA system. Although the exercises or at least some of them cannot be shown totally right (in terms of the physical movements) the participants stated, that it would be good to train with the robot, as they would easily remember the correct movements by themselves and would be motivated to make the training together with their little trainer.

In conclusion the SAR alone was preferred for motivating the trial participants to go on with the training movements during the training session in a better way than it was achieved with the video beaming approach of showing the movements. The video picture was preferred for an initial description or in detail description of different exercises. However it was also stated by the trial participants, that it could be quite hard to find a place for projecting the content due to the lack of enough space on the wall in their own homes.

Thus in future a solution including a combination of the two approaches seems to be the optimal way, by having the KSERA system as trigger system, that shows different exercises on a visual display (beaming picture, touch-screen solution, mobile screen solution,…) and in addition have NAO as motivator – on the one hand to watch the video on the different solutions, on the other hand to perform the training tasks together with the users as a kind of personal training agent.

Figure 27 shows the results of the cluster scores concerning the system compared to the State of the Art (SOTA) in home task oriented training and point out the very positive feedback concerning this comparison.

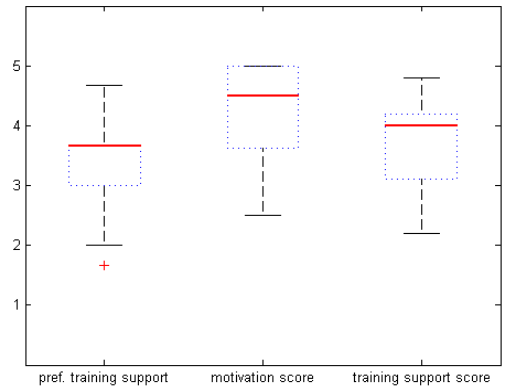


Figure 27: Score of the training support performance concerning the KSERA approach and the system (maximum value = 5, n= 32)

The very high motivation score is very promising and is a clear hint for a future application for older people in general, and especially for people with COPD, who need regular training of different parts of the body. A number of qualitative quotes by the trial participants can be found in the annex “Qualitative user comments”

3.6. Results of the Human Robot Interaction – questionnaire survey

In the following the results of the human-robot interaction questionnaire (HRI) are presented, focusing on the individual perception of the robotic system. The main aim of this questionnaire based study was to evaluate the system and generate findings about the interaction with the KSERA robotic system. These studies are interdisciplinary endeavors since they include both technical and social aspects of interaction. In this context one of the big challenges was to cope with the multitude of data acquired in user studies and to focus on the question "how can data from the user and the system be more strongly integrated".

As described above a KSERA HRI scale and a Godspeed scale were used to gather data about the subjective perception of the trial participants. In addition to the quantitative analysis a discussion about the correlation of this data with the technical performance data is presented, which is the base for future refinements and redesign of the system.

3.6.1. KSERA HRI scale – characteristics of the robot

In the following the results concerning the perception of characteristics of the robot are presented. The questions are focused on the subjective perception of robot functions, not at the body structure or appearance. These questions could then be clustered to scores describing the perception concerning

- humanoid appearance,
- movement abilities of the robot and
- conversational abilities.

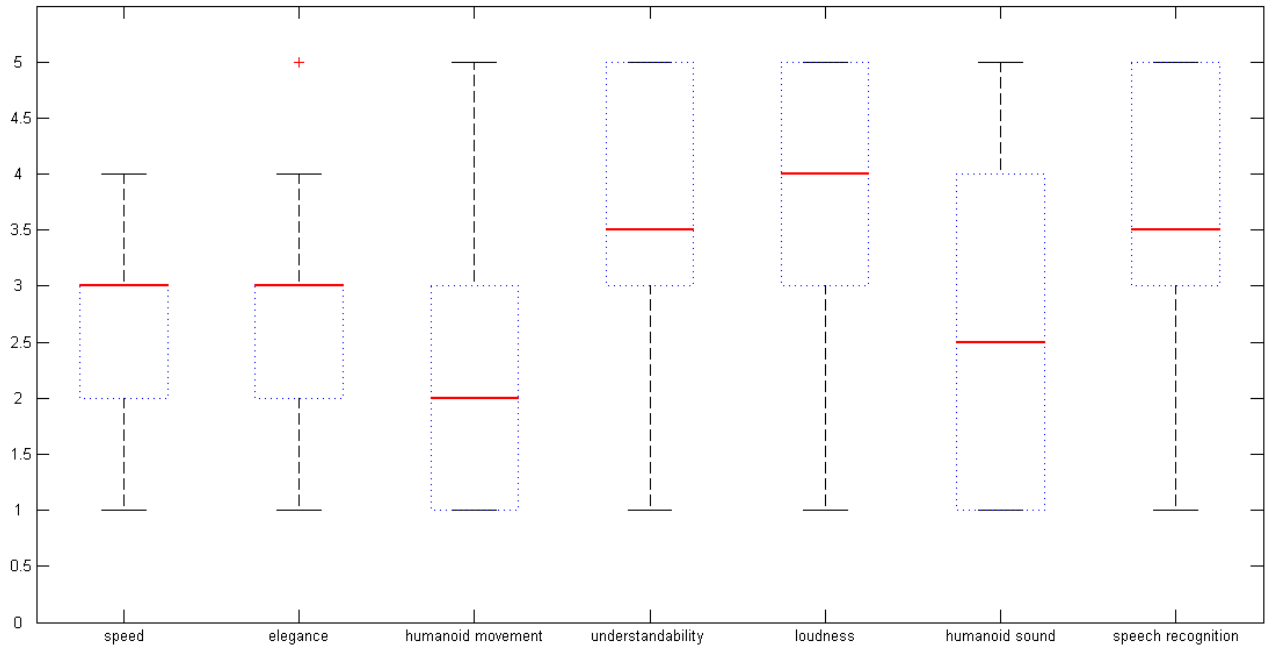


Figure 28: Scores concerning the characteristics of the robotic system

As shown in Figure 29 the perception of the **conversational abilities** was ranked quite high. This was caused by the very good speech output and understandability of the robot and on the other hand influenced by the test design. As already discussed in the technical performance chapter, the speech recognition did not work sufficiently during the trials (from the technical point of view) due to the very complex sound situation in real life environments. However the trial participants did not recognize these problems, as the tests were designed in a way that "forced" or "guided" the participants to say the right (expected) words and sentences

to NAO and thus thought the robot would understand them. This test design was used in order to conduct useful trials and give the users a good chance to see the KSERA abilities although some modules did not work in the intended way.

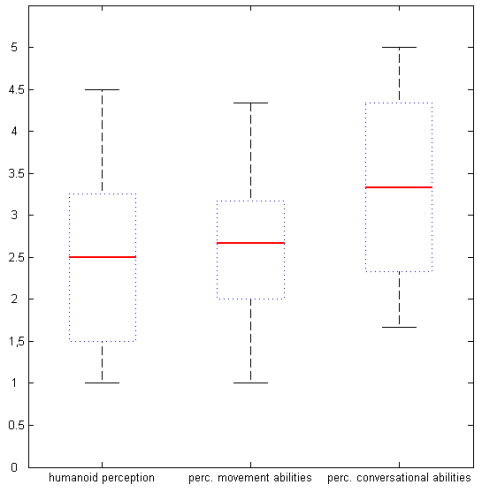


Figure 29: scores concerning the humanoid perception, perception of movments abilities and conversational abilities

The statements (quantitative and qualitative) concerning the perception of the movement abilities were also expected in the context of the performance evaluation results. Almost all trial participants stated the speed of NAO walking as too slow. Also the elegance of the movements was rated as average, which means not very fast or humanlike. This fact is probably also caused by the very complex application flow and the number of different modules working together in order to realize the robot movements in real world conditions. This is a crucial point for future development in order to enhance the humanoid perception of the robotic system. A faster and more specific and target oriented movement would be very beneficial and would definitely let the robot appear more humanlike and more knowledgeable.

3.6.2. KSERA HRI scale – attitude towards the robot

The section “Attitude towards the robot” and the associated questions deal with the overall perception of politeness and the general attitude towards the robot and how the robot’s intelligence, its ability to be knowledgeable was perceived.

The overall score for the attitude towards the robot as well as the perceived politeness was ranked high (median 3.75 of 5), which means that the robot was perceived as a friendly and polite system.

Figure 30 represents the results in a more detailed way and shows that especially the perceived friendliness of the system is ranked very high (which is a very important factor for a future long term human-robot-interaction in the homes of the users). This parameter is of course influenced on the one hand by the construction and appearance of the used SAR – NAO, on the other hand by the distinct planning of dialogs (e.g. in the German speaking version the robot used the non-formal “DU” for talking to the person) and the incorporation of factors, which are important in daily life – e.g. movement -distances to the user.

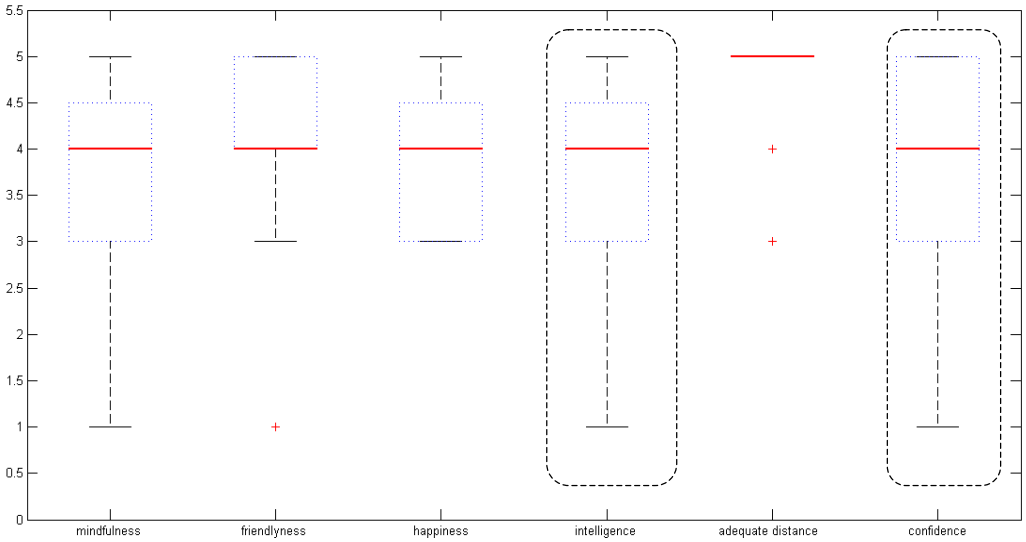


Figure 30: Scores concerning the attitude towards the robot

The robot was perceived as a mindful, friendly and happy companion, which are three important factors for motivation in daily life in several situations (disease self management, training support, reminding, warning, etc.) and are also relevant for the perception of politeness.

The intelligence of the system and especially the robot was rated above average, which is an important factor and base for the perceived confidence towards the system, which is crucial for a real benefit in daily life.

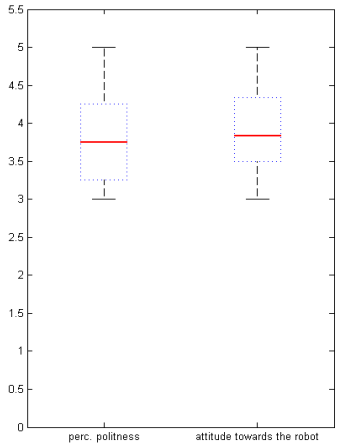
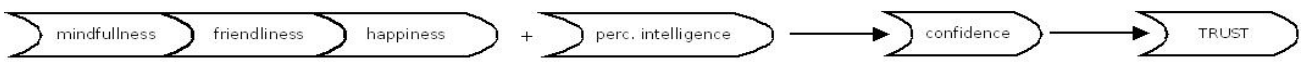


Figure 31: scores concerning perceived politeness and attitude towards the robot (cluster score)

Given these results and the experiences gathered during the field trials of PT1 the following hypothesis was constructed:

If the system is perceived as "friendly", "mindful" and "happy" the personal attitude of the user towards the system is positive. If this is the case and the system is also perceived as "intelligent" and "reliable" and the user gains confidence towards the system, which implies trusting the system.



That means that the combination of a mindful/friendly/happy robot companion, which is perceived as an intelligent companion, would have a direct effect on the perceived confidence and trust in the system. The different factors in this hypothesis are of course influenced by many others (also above mentioned factors).

Thus in order to reach the aim of optimizing the factor of confidence and perceived intelligence the related factors also have to be optimized in parallel, respectively before (as a base for them).

In this context the following factors concerning the behavior of the robot can be interpreted and have to be a base for future system enhancements. As can be seen in Figure 32 the robot's naturalness and the perceived animation were rated only on average – which is most probably correlated to the average movement and humanoid perception. For example during many trials the robot was moving correctly, but it was too time consuming to move around and place itself in the correct position.

Another problem that amplified this perception were the quite short sequences of robot interaction. Thus it was always clear for the users, that the researcher team (2 persons steering the system) would start the system. Such problems / limitations had a major effect on the perceived animation.

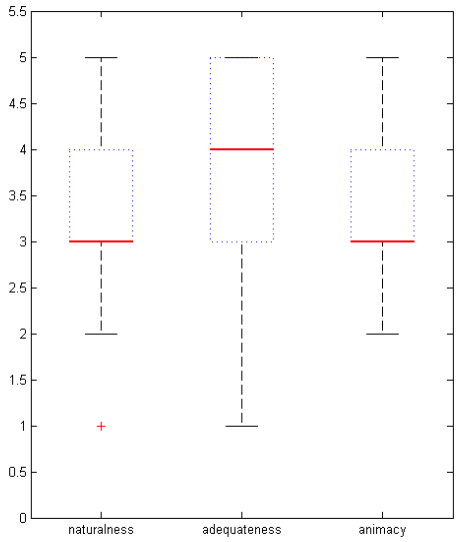


Figure 32: Scores concerning naturalness, adequateness and animacy of the SAR

In this context the perceived "naturalness" can be interpreted as well. Especially in the qualitative interviews and quotes from the users, Naos movements have been described as very machinelike (e.g. head movements, arm movements, etc.) were rated on or below the average. In the next iteration system improvements enhancing natural movements or changes (also minor movements without any clear functional purpose e.g. changes in the eyes color to simulate glimpsing or different emotions) would be very beneficial for the optimization of the perception of naturalness and animation, which are consecutively related to the perceived intelligence and perception of humanoid abilities.

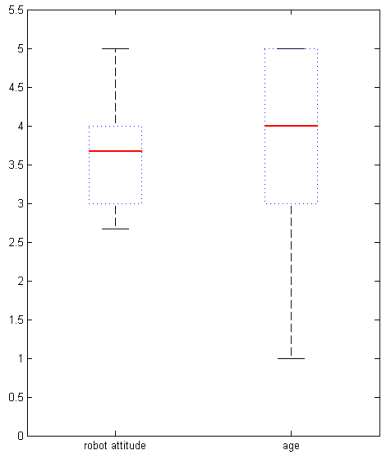


Figure 33: robot attitude and perceived age of the robot

As shown in Figure 33 the perceived age or “degree of maturity” shows that the robot is perceived between an adolescent and adult (1 – infant, 5 – adult). This is a quite interesting result, as during the qualitative interviews almost every user (the female ones even more) treated the robot more like a child and also gave a lot of quotes concerning the nice and cute appearance of the robot, which is definitely caused by NAO’s appearance. However after the test the robot was rated with a “higher” age factor, which is probably due to the perceived intelligence. Someone, who can give such distinct hints and medical parameter information, can probably not be seen as a child. Unfortunately the question for the perceived age was not asked before the trials only with the knowledge of the NAO appearance. This would be an interesting and most probably also influencing factor for perceived intelligence, and has to be evaluated in the KSERA end evaluation.

3.6.3. Godspeed Questionnaire evaluation

In the following the results of the evaluation using the Godspeed questionnaire are presented and discussed. Some of them are also related and directly comparable to the above KSERA HRI scale parameters. The Godspeed Questionnaire (Bartneck2006) is a widely used evaluation questionnaire and thus was used to guarantee comparability with related projects.

Figure 34 shows the scores for the different Godspeed clusters, which approve the above mentioned results and parameter clusters. The score for the perception of sympathy is rated very high, which shows the same result as in the KSERA HRI scale. Also the score for individual safety is very high, which is caused by the quite small appearance of the NAO robot, which has the effect, that it is not perceived as a machine that can harm anyone (this was also stated in qualitative answers during and after the test). This fact was very appreciated, even if this causes the problem, that NAO is not able to bring / carry things.

The perceived intelligence was rated comparable to the KSERA HRI scale at a relatively high level. This score and the score for individual safety also relate to the KSERA HRI scale score of perceived confidence towards the system and thus have an effect on future quality of life (QoL) questions and hypotheses.

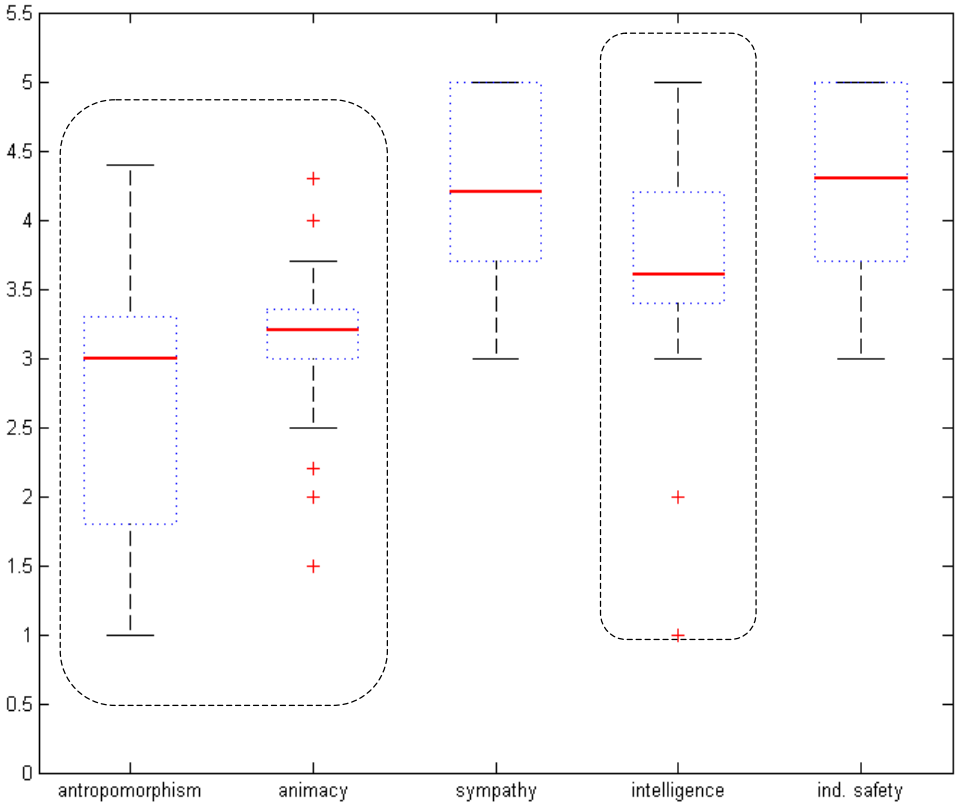


Figure 34: Godspeed clusters

Similar to the factors in the KSERA HRI scale for humanoid movements and perception of animation, the Godspeed scores for anthropomorphism and animacy are still not satisfying and were (as described above) influenced by technical performance factors like problems in the navigation or lack of interactive elements, like sub-movements without functional purposes. Those two factors are also very important for the overall perception of the robot's attitudes and intelligence. They have a clear effect on the perception of the users concerning the ability of the robot of being a helpful, knowledgeable companion in daily life, whom they can trust. Thus those parameters will be a clear indicator for the enhancement of parts and the fully integrated KSERA system over time.

A more detailed discussion and correlation discussion concerning HRI / QoL and performance test results can be found in the following chapter, which can be seen as "lessons learned description" (also beneficial for related projects) and a guide for the next KSERA development and research iterations.

3.7. Summary KSERA QoL abilities and HRI evaluation

The overall impression concerning the KSERA system and the related approach of using a humanoid robot as motivator and helper in daily life was **verified and appreciated very much by the trial** participants. The users really appreciated the several possibilities of the system also in comparison with state of the art systems (e.g. in physical exercising, reminding, motivating and physical parameter measurement). The robot NAO was perceived as a very sympathetic and safe robot, which could be used at the trials participants' homes without fear of being unsafe or being monitored during the daily life and thus can be seen as a trusted system. The fast interaction (e.g. during the task oriented training) was very good and gave a clear picture of how the future interaction (HRI) can be enhanced in real life conditions, if the interaction flows are expanded in the different testcases.

Due to the test setting and procedure it was possible to perform usability tests although parts of the system did not work as expected. However there were some errors e.g. systematic errors (such as attracting the user's attention at the wrong time) and errors driven by external conditions such as light, motion in the room (navigation) and external sound sources (speech recognition), that definitely influenced the perceived KSERA QoL and in special the HRI evaluation results. Solving these limitations has to be the base for further enhancements of the system in the KSERA prototype 2.

The movement and navigation of the robot was mostly described as too slow and not very smooth or humanoid. Also the parameter of perceived intelligence should definitely be raised by avoiding systematic errors e.g. when the robot stood in front of a person, but could not recognize her/him for a longer time, probably due to the person wearing glasses. These problems also are strongly related to the results concerning anthropomorphism and "animacy", as well as the perceived naturalness.

Another area for future enhancements has to be the research on the several interaction-input channels, as they are the base for complex interaction flows and cues. In prototype 1 a very straight forward approach concerning the human robot communication was used (in order to show the technical feasibility). The interaction and communication cues were limited to a minimum, in order to be able to perform a solid technical performance evaluation.

Most users expected to be able to talk to the robot as this felt like the most natural way of communication for them. The PT1 system included word recognition, however during the trials it turned out, that the used approach – concerning the integration of the used microphones – offered an insufficient performance rate due to real life constraints (environmental sounds, distance, etc.). In case the participant clearly observed that the robot has trouble understanding, also gestures were used. Because of the human like shape of the robot many users clearly expected to be able to communicate with the system in the same way as with a human being using voice and gestures. Thus the interaction flows and input channels have to be revised and enhanced for the KSERA end evaluation (especially the word recognition performance has to be raised) in order to perform even more realistic testcases – which will aim at longer interaction periods during the trials.

3.8. Comparison between robot system and touch screen solution

To compare the performance of the KSERA system against other state of the art systems, a local touch screen based user interface which is the outcome of the EU FP7 funded project „long lasting memories“, was used.

The final questionnaire also contained the questions

- “What system seems to be more motivating for you to really perform the desired/asked task?”
- “What seems to be preferable for you in terms of steering/choosing functions?”
- “Which system seems to be attract your attention in a better way?”
- “Why did this system attract your attention in a better way?”
- “Which system seems to be preferable for getting information faster?”
- “Which system seems to be preferable for getting better/deeper information?”
- “Which system would you prefer for explaining exercises?”
- “Why would you prefer this system?”
- “Which system seemed to motivate you the most?”
- “If you had the choice of taking one system to your home for one month – which one would you choose?”

to compare the KSERA system with a touch screen solution, which is capable of doing basically the same things as the KSERA solution does.

In most of the points there was a definite decision into one direction, which system is preferred for a certain feature. Five out of eight participants preferred when Nao was performing the exercises. The majority also said that Nao is more motivating than the touch screen – both for exercise performance and asking to do other tasks like doing measurement with the pulsoximeter and is more likely to attract attention than the screen.

The feedback about information presentation was contrary. Seven out of eight participants preferred to get information by using the touch screen. Most said that this was because they have time to read through the information on the screen and this way of presenting makes information easier to understand. Some also stated that they prefer it because they are used to get information themselves when they need it (either from teletext on tv or their computer).

Four participants stated that if they could choose, which system to take home, they would prefer Nao, two would prefer the touch screen solution and two could not decide or would take both.

The table below sums up statements by the test participants comparing the KSERA system with the used touch screen solution:

Benefits of robot system	**Benefits of touch screen solution**
• robot can move • robot speaks to user directly • pleasant, cheering up • if you are not capable of using the touch screen device anymore • robot performs exercises like trainer in	• The touchscreen PC is more practical, because you can use the PC also for other applications. • A pc is something a lot of people already have at home • It is possible to do both - hear and read information

training lesson • robot can walk through the apartment and find the user if he needs help • robot is funnier than pc • information more accurate for personal surroundings • you don't have to remember how the system works • easier to establish a relationship with the robot • exercise performance by robot makes the exercises easier to understand • more personal • user feels addressed more directly • gets more attraction when the robot moves • motivates more to do exercises than pc • robot set more actions to get attention like waving with his arms or walking to the user	• Needs certain space to place it • Alarm sound perceived as more raising attention • Movements are detected during exercise performance • User can look up information by him/herself – keeps him/her active • Information / reminder given by the touch screen has to be confirmed by the user • Signal during exercise performance is more prompting than Nao just talking • Has a fixed position → no fear of falling over the device • More time to read information • Requires a certain degree of autonomy • Easier to decide when the user wants to get what kind of information • People are more used to this kind of technology
Drawbacks of robot system	**Drawbacks of touch screen solution**
• robots distracts easier (rather looks at him than listen to information) • is not taken as serious as written info on touch screen • user does not always understand what the robot is saying	• screen has to be cleaned quite often (fingerprints) • instructions for exercises have to be read (especially bad when user has bad eye sight) • static device where the user has to walk to in order to use it • you have to remember how the system works • the robot is present – the voice from the touch screen could also come from the radio • more likely to turn off touch screen than robot • user has to read and press buttons during exercise performance • reading information can be strenuous

Additional user comments:

“If he shows the exercises right, then I would prefer the robot”

“I would have to use both systems for a longer time to decide which one I would prefer”

“A remote control for the touch screen would be nice”

“I would decide on my daily condition which system I would use – that’s more diversified”

“If I would be bedridden, I would prefer the robot, but at the moment I’m still autonomous.”

“It would be good to have it integrated in my tv instead of having an additional device.”

“Depends if you are a “visual” person or not.”

“As long as I can use it by myself and I am active I would choose the touchscreen; the robot is rather for people with difficulties”

“I just like the robot.”

“I think people will talk to the robot – even though he does not answer.”

4. Summary and Conclusions

The following table gives an overview of lessons learned during the formative evaluation, including a brief discussion on the different topics and outlook on the future prototypes in the KSERA project. Some of the points are KSERA specific, but the included content is summarized in order to give input for related projects. The results from the QoL and HRI evaluation study is therefore combined with common known facts and the overall KSERA performance. The table has a strong focus on the SAR and related performance parameters.

Field of interest	Facts in HRI	KSERA performance	Discussion on functional and technological enhancements for future prototypes and trials
Static Appearance of the SAR	• *Other studies (e.g. Duffy 2003, MacDorman2006, Waigner2006) show, that physical (and in particular humanoid) robots are preferred over virtual agents and text descriptions* • *The appearance should match the task and role of the robot in daily life (Waigner2006)* • *The appearance is not as much important as the function*	• *The evaluation showed, that the humanoid robot was seen as nice little helper, which was also trusted. The appearance was evaluated very positively.* • *The robotic approach (e.g. for exercise instructions) was seen very positively, also in comparison with text or avatar based systems. This is also related to the humanoid appearance and direct identification with the embodiment.* • *Related to the humanoid appearance of the robot, the trial participants also had the expectancy, that the robot had human abilities (e.g. understanding, communication abilities, humanoid movement – e.g. concerning speed, etc.)*	• *The function of a SAR – including the most important motivating and communicating skills – should be emphasized even more.* • *As the users expect humanoid abilities there should be a strong focus on the HRI enhancement for the future prototypes – using the expressed sympathy towards the robotic system for enhancing the user's perception of the whole KSERA approach*
Interaction and communication	• *Robots should provide humanlike communication and interaction skills – using multimodal communication possibilities (Kobayashi2008, Leite2008, Li2006).* • *The robotic (and overall) systems should also use additional channels for communication (screens, beamers, etc.)*	• *The multimodal expression abilities (talking, gestures, movements, eye-movements,…) were appreciated by the trial participants and also expected, due to the humanoid appearance of the robot. The limited number and lack of performance of the input channels (e.g. word recognition, face detection, etc.) was obvious to the participants and also recognized as a lack of*	• *The input channel abilities (in particular efficiency and effectiveness of the word recognition) have to be enhanced in order to realize longer, more realistic and adaptable communication flows and interaction cues. This will lead to a system, that will be perceived as more humanoid, which has clear effects on a possible future use in daily life. The advance has to be measured using related metrics in the KSERA HRI scale*

		the system. • *The advantage of different approaches (e.g. static monitor, which is triggered by the system) was emphasized by the trial participants. The robot should however be the central guide and motivator to use the other input/output channels*	*and Godspeed scale (e.g. perceived intelligence and animacy).* • *The abilities of NAO being a helpful motivator should be emphasized for future prototypes. This implements enhanced demonstration of (non-functional) humanlike gestures (e.g. hand-movements) and mimic-like expressions (e.g. simulated glimpsing, etc.)* • *If adequate and robust working capabilities are not possible to implement, the actions and action flow of the system should be designed in a way to match the situation (e.g. only one answer is possible, or questions will be repeated using the same question statement explained with different sentences) or even more restrict the context of use (e.g. have a clear flow of warnings/ hints without the need of user input in critical situations)* • *Express more what will happen – the robot has to be more communicative – It will be of great importance to implement clear feedback loops concerning the status and next steps of interaction – again these can be expressed by voice output, gestures or mimics. It has to be clear to the user, what the robot is performing and what is the robot's intention.* • *Provide cues for long-term (or longer scenario-) tests.*
Movement and Navigation	• *Robots should provide humanlike navigation strategies and movement patterns.* • *Robots should be fast enough to fulfill their purpose in time*	• *The possibility of the robot to move towards the user was appreciated and also wished to be performed during emergency situations* • *The KSERA system had to deal with a couple of limitations in the real life situation (lights, shades, movements of second user, etc.). This sometimes had strong effects on the navigation performance – which were also obvious to the user. During navigation tasks with only one user in the room the performance reached average values.* • *The speed of the robot and the time for the*	• *The stability concerning user and robot position detection has to be enhanced in order to have a stable base for robot navigation towards different places (in front of the user, in front of a wall, in front of different objects).* • *Monomodal sensor approaches should be replaced / enlarged by multimodal approaches, in order to meet real life requirements (e.g. different sensor information is needed in different rooms or parts of rooms f.ex. if a ceiling camera picture is not available of if sonar sensor information is misleading due to obstacles, corners or wall and floor characteristics.*

		different movement distances varied too much (however e.g. the step speed is mostly depended on the robotic platform) and the speed was evaluated as too slow.	• *Speed and movement naturalness could be enhanced in order to enhance score of animacy and humanoid perception. Those parameters will have a direct correlation to the overall acceptance of the system (Duffy 2003).* • *Cues for longer scenarios (error resistant and more adaptive cues) have to be implemented in order to meet real life requirements.* • *The SAR has to perform movements and navigation in humanlike way (concerning direction, speed and movement performance). This performance will have a direct effect on the perceived system abilities and perceived animacy and intelligence of the system.*
Support and Social Skills	• *Robots should provide humanlike skills in order to give multimodal support (give oral hints, walk to areas of interest, show movements, etc.) (Andre2004a, Andre2004b, Duffy 2003, Li2006)* • *Robots should be fast enough, to fulfill their purpose in time* • *Video communication possibilities are an enhancement in older users' daily life and offer a good opportunity in closing the digital divide.*	• *The KSERA system approach of giving hints (e.g. for monitoring special physiological parameters) or warnings was very appreciated during the trials. Also the reminding possibilities.* • *The support hints were to complex and long for the users, they would have liked shorter answers and hints, what they should – knowing the facts (e.g. air pollution – is it possible to go for a walk?)* • *The time of reaction of the system was adequate for the tested use cases.*	• *Appropriate conversational, multimodal strategies have to be emphasized.* • *Adaption towards the users' skills have to be implemented and will be an enhancement in the perception of the system's abilities.* • *Video communication has to be part of the future prototypes. Different strategies (e.g. mobile solutions using the beamer pack on the robot, using static touch screen solutions as used in smart homes and alternatives) can be used, deploying the SAR's abilities as a mobile interface and motivator, tutor, etc.*
Motivation Skills	• *A humanoid robot is a powerful tool for performing motivational tasks – deploying humanlike strategies (Kobayashi2008, Kozima2003, Leite2008, Li2006, Sidner2004).* • *Multimodal approaches show a very good performance in motivating older people for different ADLs (e.g as reminder, tutor, trainer,etc.) (Press2004, Johnson2004, Leite2008).*	• *The KSERA abilities for motivating to perform different tasks (physical parameter monitoring for self disease management, training exercise performance, critical dealing with situations) showed very good evaluation results.* • *The trial participants worked together with the robotic system and the defined aims (e.g. task oriented training, measurements, etc.) were performed by the users.* • *The trial participants gave a very positive feedback concerning the perceived intelligence*	• *Above described factors of interaction could enhance the motivational performance in future prototypes.* • *The direct activation by a mobile agent, and a humanoid SAR in special enhances and extends common known motivation strategies (verbal, avatar-based, film, etc.) (Press2004, Johnson2004, Looije2006)*

		and politeness, the sympathy towards their robotic companion and the trust towards the system.	
High Level Skills **(trust, empathy)**	• *Robots (in particular humanoid, social agents) should trigger the right user behaviors by raising trust and empathy* • *Special dialog strategies lead to optimized task performance flows (Kobayashi2008, Leite2008)*	• *The robot was seen as a sympathetic motivator and helper, whom the trial participants would trust in daily life.* • *During this test phase different dialog strategies were not yet implemented*	• *Above described factors of interaction could enhance the score for high level skills in future prototypes. A special focus should be set on longer interaction and communication cues (Lim2009, Looije2009, Looije2006, Rumpler2009), in order to improve the perceived animacy intelligence and anthropomorphism.*
Personalisation and Adaption	• *Robots will be very personal , long term companions in future, therefore strategies for maintaining relationships have to be implemented*	• *KSERA does not provide such strategies by now*	

Table 1: Discussion on achieved results and outlook

5. Literature references

[Andre2004a] *E. André, L. Dybkjær, W. Minker, and P. Heisterkamp (eds.), Affective Dialogue Systems, Tutorial and Research Workshop (ADS 2004), KlosterIrsee, Germany, June 14-16, 2004, Proceedings, Lecture Notes in Computer Science, Volume 3068, Springer Berlin Heidelbery, 2004.*

[Andre2004b] *E. André, M. Rehm, W. Minker, and D. Bühler, Endowing Spoken Language Dialogue Systems with Emotional Intelligence, in (André et al. 2004), pp. 178-187, 2004.*

[Bartneck2003] *C. Bartneck, Interacting with an Embodied Emotional Character, in J. Forlizzi, and B. Hanington (eds.), Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, ACM Press, New York, NY, pp. 55-60, 2003.*

[Bartneck2006] *C. Bartneck,* Measuring the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, 2006

[Brickmore2005] *T. W. Bickmore and R. W. Picard, Establishing and Maintaining Long-Term Human-Computer Relationships, ACM Transactions on Computer Human Interaction, 59(1):21-30, 2005.*

[Castellano2009] *G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. W. McOwan, It's All in the Game: Towards an Affect Sensitive and Context Aware Game Companion, in (Cohn et al., 2009) pp. 29-36, 2009.*

[Duffy 2003] *B. R. Duffy, Anthropomorphism and the social robot, Robotics and Autonomous Systems, 42(3-4):177-190, 2003.*

[Press2004] *Press, Menlo Park, CA, pp. 36-43, 2004. J. Grolleman, B. van Dijk, A. Nijholt, and A. van Emst, Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation, in IJseelsteijn W. et al. (eds.)*

[Johnson2004] *D. Johnson, J. Gardner, and J. Wiles, Experience as a moderator of the media equation: the impact of flattery and praise. International Journal of Human-Computer Studies 61(3):237-258, 2004.*

[Koay2007] *K. L. Koay., D. S. Syrdal, M. L. Walters, and K. Dautenhahn, Living with robots: Investigating the Habituation Effect in Participant's preferences during a longitudinal human-robot interaction study. Proceedings of the 16th IEEE International Conference on Robot and Human InteractiveCommunication (RO-MAN 2007), IEEE Press, Washington, D.C., pp. 564-569, 2007.*

[Kobayashi2008] *K. Kobayashi, K. Funakoshi, S. Yamada, M. Nakano, Y. Kitamura, and H. Tsujino, Smoothing humanrobot speech interaction with blinking-light expressions, in Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2008), IEEE Press, Washington, D.C., pp. 47-52, 2008.*

[Kozima2003] *H. Kozima, C. Nakagawa, and H. Yano, Attention coupling as a prerequisite for social interaction, in Proc of the 12th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN 2003), IEEE Press, Washington, D.C., pp. 109-114, 2003.*

[Leite2008] *Leite, A. Pereira, C. Martinho, and A. Paiva, Are emotional robots more fun to play with? in Proc. of the17th IEEE International Sympoium on Robot and Human Interactive Communication (RO-MAN 2008), IEEE Press, Washington, D.C., pp. 77-82, 2008.*

[Li2006] *S. Li, B. Wrede, and G. Sagerer, A dialog system for comparative user studies on robot verbal behavior, in Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006), IEEE Press, Washington, D.C., pp. 129-134, 2006.*

[Lim2009] *M. Y. Lim, R. Aylett, W. C. Ho, S. Enz, and P. Vargas, A Socially-Aware Memory for Companion Agents, in (Ruttkay et al. 2009), pp. 20-26, 2009.*

[Looije2006] *R. Looije, F. Cnossen, and M. A. Neerincx, Incorporating guidelines for health assistance into a socially intelligent robot, in Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006), IEEE Press, Washington, D.C., pp. 515-520, 2006.*

[Looije2009] *R. Looije, M. A. Neerincx, and F. Cnossen, Persuasive robotic assistant for health self-management of FP7-231868 SERA deliverable 3.1 Version 4/05/2010 page 21 of 22 older adults: Design and evaluation of social behaviors, International Journal of Human-Computer Studies, vol. in Press, Corrected Proof, 2009.*

[MacDorman2006] *K. F. MacDorman, Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley, in Proceedings of the ICCS/CogSci-2006 Long Symposium:Toward Social Mechanisms of Android Science, pp. 26–29, 2006.*

[Rumpler2009] *M. Rumpler, EXSTASIS – An Extended Status Model for Social Interactions, in (Ruttkay et al., 2009), pp. 49-55, 2009.*

[Sidner2004] *C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, Where to look: a study of human-robot engagement, in Proceedings of the 9th international conference on Intelligent user interfaces (IUI'04), ACM Press, FP7-231868 SERA deliverable 3.1 Version 4/05/2010 page 22 of 22 New York, NY, USA, pp. 78-84, 2004.*

[Waigner2006] *J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, The role of physical embodiment in humanrobot interaction, in Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006), IEEE Press, Washington, D.C., pp. 117-122, 2006.*

[Walters2008] *M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, K. L. Koay, Avoiding the uncanny valley: robot appearance, personality and consistency of behaviour in an attention-seeking home scenario for a robot companion. Autonomous Robots 24(2):159-178, 2008.*

6. Annex – Qualitative user comments

6.1 Austria, 1st iteration

before trials:

„he looks cute“

„does he also do vacuum cleaning? – then I would take him home with me right away”

“very cute”

“he is really sweet”

TC1:

„is he searching?“

„he goes into the wrong direction. He’s looking at me, but...“

„Hello! Are you looking for me? I am here!“ (Nao is not facing the user directly)

“Wrong direction, little boy!”

“Oh, now his eyes are gone.”

“he walks so cute”

TC2:

„Now it is squinting“

„he could show me more (different) exercises“

„I’m a bit too slow for that (arms to front and back)“

„that’s for sure not bad; especially for people who only sit at home lazily“

@beamer:“he would have bad luck with that at my home – he won’t find any space there“

„it is too bright in the room to see the video in a good quality“

„better than a personal trainer, because I do not have to leave my home“

„video might be better, because there are other people performing the exercises“

“this is cute; this would be a nice Christmas present for myself”

“he walks as clumsy as I do”

“it is slightly better when the robot shows the exercises himself as the people on the video are a bit distracting”

“come on, Susi, let’s do something.” (user is waiting for the robot to start with the exercises)

“I’m going to start doing exercises with this.”

“I would prefer the robot with beamer for showing the exercises, because it is livelier”

“this is cute” (Nao starts to walk)

“is he nervous?” (Nao is tapping from one leg to the other in front of the wall)

“I prefer Nao performing the exercises, because it is impressing and somehow funny.”

“Rather using Nao without beamer is also a question of space.”

“This is good that’s motivating me to join him”

“You better see how the exercise really works on a video with real people.”

“He does not do the exercise movements completely right”

“the version with beamer is better, because it is real people who show the exercises and you can be sure that the exercises are shown correctly”

“it would be good to get feedback and further (vocal) motivation by the robot”

“it is better when the robot shows the exercises himself – he is doing that quite naturally”

“it is easier to imitate the exercises when Nao is doing them as well than with the video image”s

TC3:

“now I’m curious, if you tell me the right things, honey”

“now he said a wrong value”

“this was already very good.”

“I can understand him very well, but I can still hear very well.”

General comments:

„I can not understand him well (pronunciation and volume)“

„if he can do more, he certainly will be able to help me in case of emergency“

„he doesn’t do any harm to me“

„if he can do more, I would use him every day. Why would have him, if I would not use him every day?“

„he does not scare me“

„I’m not surprised that something like that exists“

„I would call him a ‚good spirit’ or ‚little helper’“

„I think that this is good, I think this is very good.“

„For people who can not go outdoors or are alone.“

„if a person is in a wheelchair or is forgetful, then this is a very good thing“

„I am positively surprised“

„I expected it to be smaller“

„if I would feel bad, I would use it every day, but I’m still very active.“

„automated reminder would be better as it is very easy to lose a piece of paper.“

„you can not interpret the measured data yourself very good“

„I don’t need the robot to tell me about my physical status – I can feel that myselfwhen something is wrong“

“he has a very comfortable voice, which is easily understandable”

“he is cute, nice and very funny.”

“to help in case of emergency, refinements need to be done!”

“he awakens the child in me”

“For me he would be very useful for doing physical exercises, because he motivates me a lot to do the exercises.”

“it is very interesting, but still needs a lot of improvement for real world use.”

“as long as the data does not leave my flat I don’t feel observed.”

“for some people it might be very good as reminder”

“this is more like a game; I can not imagine that this can help anyone who is frail.”

“the different colors in his eyes are a bit scary”

“it is great, if you live alone and need some help in remembering different things like taking pills”

“I think it’s funny and I can imagine that people can establish a personal relationship to the robot”

“he is a bit like a dog”

“it is better than being alone”

“it is a boy – he already looks like one.”

“he seems to be a bit scared” (interpreting Nao’s eyes)

“he is talking very silently; (turning zu Nao:) ‘can you talk a bit louder, please?’”

“I shouldn’t look at him, otherwise I don’t really pay attention to what he is saying.”

“He is looking funny”

“the voice does not really suit him – should rather be a man’s or a child’s voice.”

“I can imagine that some people fall in love with this little guy”

“you have to get used to it”

“for active people like me it is too slow – there as to be more ‘action’, but for frail, dependent people it is ok”

“he could be more demanding in his voice or say things twice”

“you have to be fit in mind to use it”

“you get the feeling that someone looks after you, reminds you”

“I’m feeling cared for.”

“he is too slow – I have already done everything until he is here.”

“I’m not scared by technology, but it is too expensive”

“interpretion of medical data by a doctor would be better than by a robot”

6.2 Austria, 2nd iteration

TC1:

Nao starts to walk: “he is a funny guy”

“now his eyes are black” (after 1st waving)

“reminding on taking pills or drinking is important”

“first he was walking in my direction, but then he probably thought that he does not want me anymore”

TC2:

“that’s very good getting the exercises demonstrated by the robot”

After Nao says ‘thank you’: “Thank you too”

“it is good to overcome one’s weaker self”

“criticism during exercises would be good”

“it would be interesting, if he would grouch when I’m not doing the exercises – that is something that is still missing”

“he is explicitly talking to me”

TC3:

“some words are hard to understand”

“it is good to get this info”

“it would be better, if he only says ‘too high’ or ‘too low’ and what you should do” (too much and too complicated information)

General comments:

“it is funny”

“it can be very helpful for people who live alone”

“it helps staying active”

“I don’t like computers”

“I did not expect it to be that great; I rather expected it to be boring, but I’m gladly surprised”

“It is good when reprehends you”

“I believe that this will have a stake in the future”

“as long as all data stays in my flat I don’t feel monitored”

“it is surely great, when everything is working well”

“he does not walk yet where he should walk to”

“he is so cute”

“in my opinion: the idea is great”

“he would have helped my mother a lot”

“you don’t really have to do something to interact with him”

“he might become a kind of flat mate over a longer period of time”

“it would be good, if a doctor could access the data as well”

“the system tries to help the older generation and to take the feeling of loneliness; the robot is someone who is there for them solely; you can talk to him about things that you have at heart at that moment”

“it is personal although it is technical – somebody is talking to me directly”

“he is so cute”

“my mother surely would have done her physical exercises better with the robot”

“I’m convinced that people will build up a relationship with the robot – like with puppets a lot of older people play with”

“I am amazed”

“he would be very helpful, if I were forgetful.”

“it would be good, if he would admonish me to do something”