ICT-2009.7.1
KSERA Project
2010-248085

Deliverable 1.3

SAR evaluation metrics for elderly in a home environment

31 January 2012
Public Document

©KSERA consortium

Project acronym:	KSERA
Project full title:	Knowledgeable SErvice Robots for Aging
Work package:	1
Document number:	D1.3
Document title:	SAR evaluation metrics for elderly in a home environment
Version:	1

Delivery date:	January 2012
Actual publication date:	January 2012
Dissemination level:	Public
Nature:	Report
Editor(s) / lead beneficiary:	Johannes Oberzaucher (RALTEC) Franz Werner (RALTEC)
Authors(s):	Johannes Oberzaucher (RALTEC) Franz Werner (RALTEC) Hadas Lewy (MAC) Katharina Werner (RALTEC)
Reviewer(s):	Antonella Frisiello Cornelius Weber

©KSERA consortium

Glossary of terms used

Acronym/terms	Definition
ADL	Activities of Daily Living
AR	Assistive Robotics
COPD	Chronic Obstructive Pulmonary Disease
ER	Emergence Room
FEV1	Forced expiratory volume in 1 second
FVC	Forced expiratory vital capacity
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 Centred Design
Scenario	Usage narrative
Use case	Describe the functional requirements
WHO-ICF	See International Classification of Functioning
WP	Work package
TP	Trial Participant (User)
TC	Testcase
Research team	Crew that performs the tests
PEOF	Perception of ease of use
PT1	KSERA System Prototype used for the formative evaluation
PT2	KSERA System Prototype used for the end evaluation
PU	Perception of usage
QoL	Quality of life
STC	Sub-testcase – smallest part of a testcase
IP	Injection Point – a point at which the testers are able to restart the test
TS	Test subject – end user who participates at the trial

Contents

Executive summary **5**

Purpose of this deliverable **5**

Suggested readers **5**

Relationship to other documents **5**

1. Metrics used during the trials and evaluation model **6**

1.1. KSERA Evaluation Model – Mix of quantitative and qualitative methods *6*

2. (Basic) initial assessment **9**

2.1. WHO Quality of Life *9*

2.2. Attitude towards technology - Usage of technical devices in daily life *14*

2.3. PANAS *16*

2.4. CCQ – Clinical COPD questionaire (24h version) *18*

3. KSERA Performance testing metrics – Aspects of Effectiveness, Efficiency, Flexibility, Robustness, Utility **19**

4. Metrics related to Social Acceptance (User Acceptance) and Social impact (Quality of life, health and security) **25**

4.1.1. Hypthesis related to the QoL 26

4.1.2. KSERA QoL Scale 27

5. Human Robot Interaction – Perception of the SAR system **38**

5.1. KSERA HRI Scale *38*

5.2. HRI – Godspeed Questionaire *42*

6. References **45**

7. ANNEX – Analysis sheets **46**

Executive summary

This deliverable includes metrics, that are used during the evaluation phases in the project KSERA during the formative evaluation and end evaluation (performed in WP5). It includes metrics concerning basical, initial assessment of the trial participants, performance testing, user acceptance evaluations and dedicated human-robot-interaction evaluation.

It included measures to evaluate the user experience of older people using the Socially Assistive Robots (SAR) and in particular the KSERA application. The work reported in KSERA deliverable D1.3 is input to WP5.

Purpose of this deliverable

This document describes metrics that will be used during the evaluation periods. These metrics are base for evaluation of the different testcases described in the KSERA deliverable D5.1. They can be used to directly extract questionnaires (localized for every trial performing country) for the trials.

Suggested readers

This document is used by the trial performing partners (RALTEC, MACCABI) and can also be used performing lab trials for evaluating sub-components. The proposed metrics and included criterias can be used for related projects.

Relationship to other documents

This document is strongly connected to KSERA deliverable D5.1 and to the upcoming resulting deliverables in T5.3 the Formative Evaluation and T5.2 End Evaluation.

©KSERA consortium

1. Metrics used during the trials and evaluation model

The rapid prototyping in KSERA workpackages WP2, WP3 and WP4 and the prototypes (PT1, PT2) tested in the end-user tests in WP5 will use the metrics described in this document. The evaluation measures take into account high-level constructs such as **user acceptance, performance, usability, efficiency, satisfaction, and enjoyment**.

Existing tools for measuring user satisfaction and HRI performance, as well as qualitative and quantitative adapted methods will be considered. The metrics define **what** should be measured and **how** it should be measured.

For instance ISO standards on usability define efficiency (time) and effectiveness (number of errors). Measures to evaluate the user experience in HRI are the focus of this document.

The metrics which will be used during the end user evaluations will be divided in the domains

1. **(Basic) initial Assessment**
2. **KSERA performance testing metrics – task effectiveness**
3. **User acceptance testing and metrics related to QOL**
4. **User experience concerning the human robot interaction**

The basic initial Assessment metrics will be used for evaluating the preconditions concerning the trial participants status, technology knowledge and attitude towards technology. These metrics will be used once in the beginning of the trials for each trial participant.

The other three domains will be used during every test session. A more detailed description for what test-cases and in what order can be found in KSERA deliverable D5.1.

1.1. KSERA Evaluation Model – Mix of quantitative and qualitative methods

The **KSERA Evaluation Model** is focused on the research question *how users come to accept and use the system*. Beside the technical performance (but strongly related to it) the model suggests that when the trial participants are presented with this new kind of technology, a number of factors influence their decision about how and when they would use and accept the system in a future use

The main aim of the measures included in this document is to evaluate the user experience of older people using the Socially Assistive Robots (SAR) and in particular the KSERA application. Therefore a mix of different Assessments, Usability tests and Performance tests was created. The combination of the different metrics should lead to an interpretable result, concerning the overall performance of the KSERA system in respect to the user experience. Whereas this performance is strongly related to

conclusions about perceived usefulness (PU) and perceived ease-of-use (PEOU) of the system (see also [Davis1989]).

- **Perceived usefulness** (PU) – *„…the degree to which a person believes that using a particular system would enhance his or her daily performance".*
- **Perceived ease-of-use** (PEOU) "…the degree to which a person believes that using a particular system would be free from effort"

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

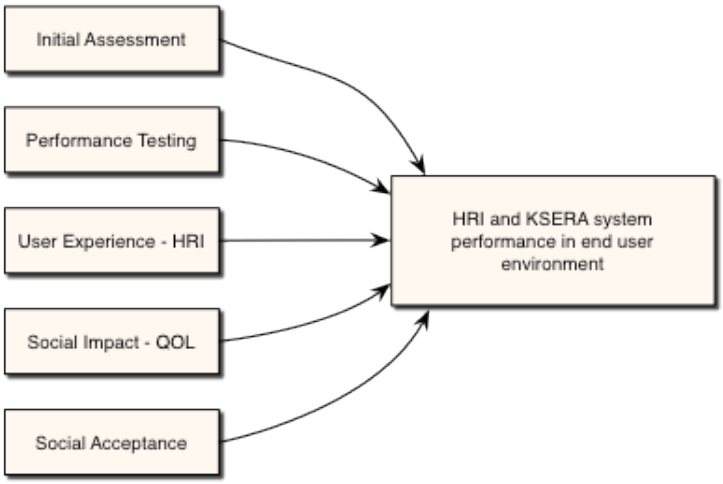

Figure 1: method mix used during the KSERA evaluation phases

Table 1 gives an overview of the different used methods in the described measuring domains. This representation shows the two big method domains. On the one hand **retrospective video analysis** – mostly used for evaluation of the KSERA system performance concerning technical research questions. On the other hand questionnaire based qualitative and quantitative measures, which will be performed during and right after the test sessions.

Research Area	Methods	Primary user studies - retrospective analysis	Expert studies	Questionaires	Interview	Focus groups
Initial Assessment						
	Analysis of current QOL			x		
	Analysis of current emotional state			x		
	Attitude towards using technology			x		
Performance Testing						
	Effectiveness	x				
	Efficiency	x				
	Flexibility	x				
	Robustness	x				
Social Acceptance						
	Performance Expectancy			x		
	Analysis of users' system acceptance			x		
Social impact - QOL						
	Analysis of support abilities -physical parameter measurement	x	x	x		x
	Analysis of support abilities - excercise	x	x	x		x
	Analysis of support abilities -entertainment	x	x	x		x
	Analysis of support abilities - emergency	x	x	x		x
User experience - HRI						
	Embodiment, robot attributes			x		
	Emotion, robot attitude			x		
	Human Oriented Perception			x		
	Co-Experience, relationship human-robot			x		x

table 1: Overview of the different testing domains and methods used

2. (Basic) initial assessment

The basic initial assessment will be performed one time in the beginning of the test sessions. The aim is to assess the current state (**WHO Quality of Life**) of the trial participants concerning their "quality of life" and their current emotional state (**PANAS**). Both will be done in order to better understand and interpret the later results and data which was gathered during the test sessions.

Additionaly a **questionnaire concerning the current usage of technical devices** in the daily life is presented, which results will represent the attitude towards using technology.

Before every test session an assessment concerning the physical status of the trial participants will be performed. A clinical COPD questionare (**CCQ**) will be used before every test-session in order to assess possible problems/extraordinary situations concerning the physical status of the trial participants.

2.1. WHO Quality of Life

Parts of this questionary [Murphy2000] are done **once at the beginning of the user trials**.

*„…The aim was to develop an **international cross-culturally comparable** quality of life assessment instrument. It assesses the individual's perceptions in the context of their culture and value systems, and their personal goals, standards and concerns….*

…The WHOQOL instruments were developed collaboratively in a number of centres worldwide, and have been widely field-tested. The WHOQOL-BREF instrument comprises 26 items, which measure the following broad domains: physical health, psychological health, social relationships, and environment…

…The WHOQOL-BREF is a shorter version of the original instrument that may be more convenient for use in large research studies or clinical trials…."

*[*http://www.who.int/substance_abuse/research_tools/whoqolbref/en/*]*

Application:

Once in the beginning of the trial period, before the trial participant has seen the system and the test-setting. The questionnaire will be filled out in a separate research room together with a member of the research team.

WHO Quality of Life QUESTIONAIRE:

Tester	Date Completed

Please read each question, assess your feelings, and circle the number on the scale that gives the best answer for you for each question.

(Please circle the number)

	Very poor	Poor	Neither poor nor good	Good	Very Good
How would you rate your quality of life?	1	2	3	4	5

(Please circle the number)

	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
How satisfied are you with your health?	1	2	3	4	5

The following questions ask about how much you have experienced certain things in the last two weeks.

(Please circle the number)

	Not at all	A little	A moderate amount	Very much	An extreme amount
To what extent do you feel that physical pain prevents you from doing what you need to do?	1	2	3	4	5
How much do you need any medical treatment to function in your daily life?	1	2	3	4	5
How much do you enjoy life?	1	2	3	4	5

(Please circle the number)

	Not at all	A little	A moderate amount	Very much	An extreme amount
To what extent do you feel your life to be meaningful?	1	2	3	4	5

(Please circle the number)

	Not at all	Slightly	A Moderate amount	Very much	Extremely
How well are you able to concentrate?	1	2	3	4	5
How safe do you feel in your daily life?	1	2	3	4	5
How healthy is your physical environment?	1	2	3	4	5

The following questions ask about how completely you experience or were able to do certain things in the last two weeks.

(Please circle the number)

	Not at all	A little	Moderately	Mostly	Completely
Do you have enough energy for everyday life?	1	2	3	4	5
Are you able to accept your bodily appearance?	1	2	3	4	5
Have you enough money to meet your needs?	1	2	3	4	5
How available to you is the information that you need in your day-to-day life?	1	2	3	4	5
To what extent do you have the opportunity for leisure activities?	1	2	3	4	5

(Please circle the number)

	Very poor	Poor	Neither poor nor well	Well	Very well
How well are you able to get around?	1	2	3	4	5

The following questions ask you to say how good or satisfied you have felt about various aspects of your life over the last two weeks.

(Please circle the number)

	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
How satisfied are you with your sleep?	1	2	3	4	5
How satisfied are you with your ability to perform your daily living activities?	1	2	3	4	5
How satisfied are you with your capacity for work?	1	2	3	4	5
How satisfied are you with your abilities?	1	2	3	4	5
How satisfied are you with your personal relationships?	1	2	3	4	5
How satisfied are you with your sex life?	1	2	3	4	5
How satisfied are you with the support you get from your friends?	1	2	3	4	5
How satisfied are you with the conditions of your living place?	1	2	3	4	5
How satisfied are you with your access to health services?	1	2	3	4	5

(Please circle the number)

	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
How satisfied are you with your mode of transportation?	1	2	3	4	5

The following question refers to how often you have felt or experienced certain things in the last two weeks.

(Please circle the number)

	Never	Seldom	Quite often	Very often	Always
How often do you have negative feelings, such as blue mood, despair, anxiety, depression?	1	2	3	4	5

Did someone help you to fill out this form? *(Please circle Yes or No)*	Yes	No

©KSERA consortium

2.2. Attitude towards technology - Usage of technical devices in daily life

The following questionnaire will be used to assess the trial participants attitude towards technology and the use of ICT and computers in their daily life. This should offer the opportunity to better understand results related to the user acceptance and HRI testing.

Application:

Once in the beginning of the trial period, before the trial participant has seen the system and the test-setting. The questionnaire will be filled out in a separate research room together with a member of the research team.

Attitude towards technology QUESTIONAIRE

1. Please indicate the technical devices you use during your daily living
   - TV
   - Mobile phone
   - Computer
   - Radio
   - Medical Reminder
   - Physical Measurement Tools
   -

Others:_____

2. Do you have an internet connection in your home
   - yes
   - no

3. What is the TOTAL number of hours a week that you spend on a computer?

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

4. What is your most common activity when using a computer? (check ONLY ONE)
   Internet surfing
   - Email
   - Games
   - Writing
   - Getting information
   - Social activities

   Others? _____

5. Do you use the internet more or less than your friends/relatives?

more less same as

6. How many hours a week do you spend doing the following activities:

Internet surfing

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

Using email

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

Searching for information in the internet

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

How many hours a week do you spend talking on the telephone?

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

How many hours a week do you spend watching television?

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

How many hours a week do you spend watching television with friends?

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

How many hours a week do you spend socializing with friends?

0	Less than 1	1-3	4-6	7-10	11-15	16-20	20+

2.3. PANAS

The panas score can be used to determine ***the current emotional state of the user*** which influences the outcome of the trials and the results concerning the perception and acceptance of the system. This questionary is done during every iteration of the user trials to obtain the current emotional state of the user.

POSITIVE AND NEGATIVE AFFECT SCHEDULE (PANAS)
20-item test using 5-point scale that ranges from very slightly (1) to extremely (5)

"In recent research, two broad, general factors—typically labeled Positive Affect (PA) and Negative Affect (NA)—have emerged reliably as the dominant dimensions of emotional experience. These factors have been identified in both intra and interindividual analyses, and they emerge consistently across diverse descriptor sets, time frames, response formats, languages, and cultures."

To measure these factors, Watson, Clark, and Tellegen [Watson1988] developed the Positive and Negative Affect Schedule (PANAS), which consists of two 10-item scales for PA and NA, respectively."

Application:

The questionnaire will be used before every iteration of the tests. The questionnaire will be filled out in a separate research room alone by the trial participant.

PANAS QUESTIONAIRE

Below there are some questions listed concerning the emotions regarding the way you are **feeling right now** .

Next to each question there is a scale describing to what degree you may/may not feel each emotion. **Please circle the number corresponding to the statement that best describes how you are feeling RIGHT NOW about the system!**

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

	1	2	3	4	5
	very slightly or not at all	a little	moderately	quite a bit	extremely

20 PANAS items

Active	_____
Enthusiastic	_____
Determined	_____
Attentive	_____
Inspired	_____
Strong	_____
Interested	_____
Alert	_____
Excited	_____
Proud	_____
Afraid	_____
Nervous	_____
Scared	_____
Upset	_____
Guilty	_____
Hostile	_____
Ashamed	_____
Jittery	_____
Irritable	_____
Distressed	_____

2.4. CCQ – Clinical COPD questionaire (24h version)

The CCQ will be used before every test-session in order to assess possible problems/extraordinary situations concerning the physical status of the trial participants. On the CCQ website [CCQ] an online form can be found, which automatically calculates a CCQ score

„…The Clinical COPD Questionnaire (CCQ) is a short health status questionnaire for patients with COPD. The main purpose of this questionnaire is the measurement of health status of patients with breathing difficulties due to their illness. By measuring health status, the general practitioner, the specialist, and the patient get better information on the severity of symptoms and limitations caused by the disease. The effect of treatment on symptoms and limitations can be measured as well…."

„…The main purpose of this questionnaire is measurement of disease related health status of the patient. The questionnaire does not intent to diagnose COPD…"

Attention: The use of the questionnaire in international research, in which different languages are used is only allowed after approval of the author.

Application:

The questionnaire will be used before every iteration of the tests. The questionnaire will be filled out in a separate research room alone by the trial participant.

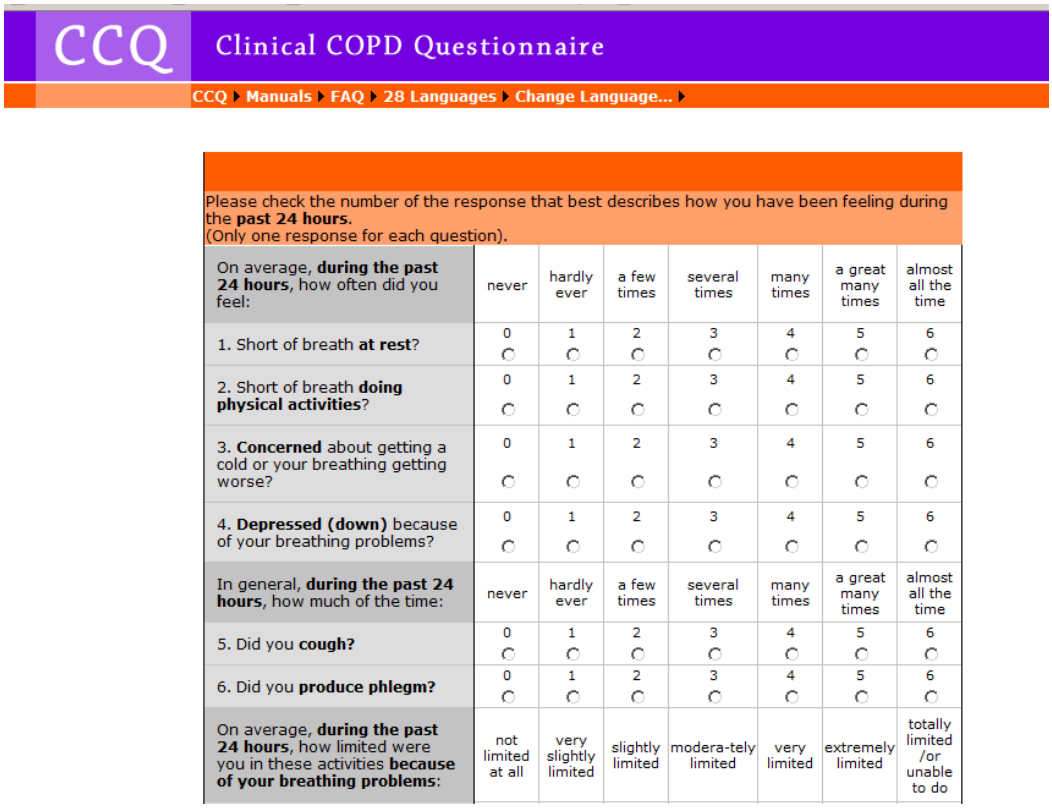

figure 2: CCQ online form

3. KSERA Performance testing metrics – Aspects of Effectiveness, Efficiency, Flexibility, Robustness, Utility

As one of the main factors for the perception of the usefulness (PU) and ease of use (PEOU) [Davis1989] the performance of the integrated system (PT1 and PT2) will be in the focus of both evaluation phases (T5.3 Formative Evaluation and T5.4 End Evaluation). Technical functionalities and related performance will be evaluated, which **are user-related**. A technical component as well as system testing has been already performed in other project phases. The following metrics focus on aspects of

- **Effectiveness,**
- **Efficiency,**
- **Flexibility,**
- **Robustness**

Description of the domains – indicated by different domain IDs:

- A_i……related to the testing of the SAR's abilities,
- B_i ……related to the testing of the ubiquitous, environmental and physical measurement components,
- C_i ……related to the testing of the KSERA system as a whole

ID	Metric name	Research question	Method	Measured- Value	Evaluation Criteria
A1	**NAO's movement rate**	Is NAO performing its' movements in the intended way?	• Visual interpretation and/or problem identification • Retrospective Analysis • System log entry	• **Bool – T/F** • Opt: Difference of optimal/intended movement and real movement – **qualitative textual description.**	• True if movement was completed as planned e.g. NAO correctly reached the desired spot / showed the exercises correctly
A2	**NAO's movement time**	How long does it take NAO to perform the intended movement?	• Visual interpretation and/or problem identification • Retrospective Analysis • System log entry	• Time to accomplish the intended movement [s] - quantitative	• Time between NAO starting the movement to the time NAO ended the movement / started

					with the next movement
A3	**User's position recognition rate**	Is the KSERA system calculating the right user position?	• Observation of the "user-area" on the ceiling camera picture and comparison of the KSERA user coordinates area • Retrospective Analysis of situation • System log entry	• End position right: Yes / No **[bool] - quantitative** • Number of mismatches over time FP **[n] - quantitative**	• True if the calculated user position is not more than 0.5m away from the actual users position during the respective testcase
A4	**Recognition of NAO's position rate**	Is the KSERA system calculating the right NAO position?	• Observation of the "NAO-area" on the ceiling camera picture and comparison of the KSERA user coordinates area • Retrospective Analysis of situation • System log entry	• Single end position right: Yes / No **[bool] - quantitative** • Number of mismatches over time FP **[n] - quantitative**	• True if the calculated robot position is not more than 0.5m away from the actual robot position during the respective testcase
A5	**Recogntion of the user's gaze direction rate**	Is the KSERA system calculating the right user's gaze direction towards NAO?	• Observation of the user's gaze direction on the ceiling camera picture and frontal camera and comparison of the KSERA gaze direction calculation • Retrospective Analysis of situation • System log entry	• Single approach during one testsession: Yes / No **[bool] - quantitative** • Number of mismatches (FP,FN) **[n] - quantitative** • Error rate (FP,FN) during one testsession **[%] - quantitative**	• True if the difference between the calculated gaze angle and the true observed angle is less than 45°
A8	**Robot speech /phrase / word recognition rate**	Is the KSERA system understanding – the user's commands	• Observation of the NAO and human user via camera system and microphone • Retrospective Analysis of situation • System log entry	• Single approach during one testsession: Yes / No **[bool] – quantitative** • Number of mismatches (FP,FN) **[n] - quantitative** • Error rate (FP,FN) during one testsession **[%] - quantitative**	• True if the KSERA system calculated the correct meaning of the users words. (e.g. "yes" in case the user said "yes please" is correct.)
A9	**Robot speech output rate**	Is Nao "saying" the expected sequence	• Observation of the NAO and human user via camera system and microphone • Retrospective Analysis of situation • System log entry	• Single approach during one testsession: Yes / No **[bool] - quantitative** • Textual descripition of the spoken sequence in case of wrong sequence	• True if the system reported the correct sentence

A10	**Robot speech understanding by user rate**	Are Nao's words understandable?	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Single approach during one testsession: Yes / No **[bool] - quantitative** • All: Error rate [%]	• False if the user complained about the understandability of if the user clearly did not understand Nao correctly • Otherwise true
A6	**Neglect Tolerance(NT)**	Measures the autonomy of the KSERA system (and especially NAO) with respect to the task Measures of how robot's effectiveness declines over time when the robot is neglected by the user	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation System log entry	• NT **[s] – quantitative** NT - Neglect tolerance is our basic mechanism for measuring the autonomy of a robot. The amount of time that a human can ignore a robot has a lot to do with the attention leverage that the robot can provide. • Measuring neglect tolerance: Amount of time that a Human can ignore a robot • *"Place the robot at a some random location in a problem world and giving it a random goal to achieve and measure the time that the robot is effective: the elapsed time during which the robot makes progress towards that goal before dropping below the effectiveness threshold"* • *"Time between some user instruction and either dropping below the effectiveness threshold or a new user instruction"*	
A7	**Robot attention demand (RAD)**	How long is the user's time consumed by interaction with NAO? How long does the user have to perform action to make the system work in the intended way?	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• RAD = IE/(IE+NT) - **Uniteless quantity – quantitative** IE - is the amount of time required to interact with the robot. Examples: • The teleoperated robots have a small NT and RAD approaches to 1: the user can focus on other things besides interacting	

				with the robot • Reducing RAD can be done by increasing NT or decreasing IE but not always (NT ang IE are NOT independent)	
A11	**Robot getting attention cycle (only for PT2)**	How long does it take to get the user's attention	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Time by attention cycle [s]	
A12	**Human user manipulation/motivation by the robot rate**	Is the KSERA system (and especially NAO) motivating the user to perform special tasks	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Single approach during one testsession: Yes / No **[bool] - quantitative** • All: Motivation rate TP [%]	• True if the user undertakes the task it was motivated to do.
A13	**Video display rate**	is the video presented in a way useful for the user (size, quality, readability,...)	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No during one test session [bool] • Nr.of mismatches – error rate [%]	• True if the video was displayed as a whole and the quality was approved by the user
A14	**Music output rate (only for PT2)**	is the music played in an appropriate way (quality, volume, timing,...)	• Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • Textual description • Nr.of mismatches – error rate [%]	• True if the music was played correctly as a whole and the quality was approved by the user
A15	**KSERA Face detection rate**	Did the KSERA system correctly find the users face	• Observation of the KSERA tester interface • Observation of the NAO and human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • Textual description • Nr.of mismatches – error rate [%]	• True if the users face is recognized and displayed in the tester interface at the right time
A16	**Face detection duration**	How long did it take the system	• Observation of the scene via the camera system and the KSERA	• Time to accomplish the face detection . quantitative in seconds	• Time from beginning the face detection movements (turning

		to find the users face?	tester interface • Retrospective analysis of the recorded material and the system logs		the head in order to get a proper view of the user) until the users face was found (A15)
B1	Physical Parameter Measurement	Does the KSERA system trigger the data	• Observation of the human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • success rate [%] • Error rate [%]	• True if the data could be gathered by the system correctly (e.g. no missing values, correct readings)
B2	Physical Parameter Measurement	How long does it take to trigger the phys. measurement data	• Observation of the human user via camera system • Retrospective Analysis of situation • System log entry	• Time from starting the measurement • To data base entry [s]	• Time taken in seconds between the user undertaking the measurement and the systems log entry in the database
B3	Actuation (PT2)	Does the KSERA system perform the actuation	• Observation of the human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • success rate [%] • Error rate [%]	• True if the actuation was triggered correctly
B4	**Data propagation**	Is the system able to store gathered data, retrieve it, act upon it	• Observation of the human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • success rate [%] • Error rate [%]	• True if the system was able to react according to retrieved external data (e.g. physical / environmental measurements)
B5	**Environmental parameter rate**	Is the system able to correctly gather env. indoor and outdoor parameters	• Observation of the human user via camera system • Retrospective Analysis of situation • System log entry	• Single: Yes / No [bool] • success rate [%] • Error rate [%]	• True if the data was gathered correctly and is available for analysis and reactions of the system
				•	•

C2	Timing rate	Does the KSERA system perform timing operations correctly (E.g. wait for user input, ...)	1. Observation of the NAO and human user via camera system 2. Retrospective Analysis of situation 3. System log entry 4. Time measurement	• Single: Yes / No [bool] • Time for task [s]	• True if the overall testflow showed no timing difficulties such as major delays between tasks.
C3	Task sequence rate	Is KSERA able to perform the correct sequence of tasks in the right order?	1. Observation of the NAO and human user via camera system 2. Retrospective Analysis of situation 3. System log entry	• Single: Yes / No [bool] • Textual description	• True if the tasks were executed in the right sequence.
C4	Decision making rate	Is KSERA able to derive the correct decisions based on input data?	1. Observation of the robot and human user via camera system 2. Retrospective Analysis of situation 3. System log entry 4.	• Single approach T/F [bool] • Success rate [%]	• True if the decisions drawn are consistent with the received inputs.

4. Metrics related to Social Acceptance (User Acceptance) and Social impact (Quality of life, health and security)

Beside the question for a socical (user) acceptance of the KSERA system, one of the main research questions in KSERA is *"Is the KSERA approach able to improve the quality of life of older persons in general and in special those with COPD?".* However the measurement of the described quality of life and all related aspects lacks of a definition of *"What is quality of life?".*

The term ***quality of life*** is used to evaluate the general well-being of individuals. The term is used in a wide range of contexts and mostly is adapted towards concrete research topics and related questions. Standard indicators of the quality of life include physical and mental health, but also aspects of safety, privacy and of course self-determination.

Concerning the evaluation of the QoL during the two evaluation phases 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

- Physical 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 trough Excercise" and "Disease Self management").

Based on these domains measurements concerning the question of the perceived usefulness of the KSERA approach regarding to the **acceptance** of the KSERA system and **a possible improvement in the quality of life (QoL)** of the later users will be covered with the following questionaires. These questionnaires cover questions concerning the approach and not concerning the technical performance, however information for improvements for PT2 can be included.

As a starting point, hypothesis (research questions) were defined, which give an overview about the general research questions during the trials. This research questions will then be answered deploying the following metrics. The information will be gathered by doing interviews, using the successive questionnaires. The questionnaires will offer the opportunity to gather qualitative as well as quantitative data.

The general topics of research questions (for the "Formative Evaluation") are divided into questions dealing with the named domains: *"General acceptance of the KSERA approach", "Physical measurement and ICT based health support abilities", "Exercise and Training support abilities" and "Disease self management"*

4.1.1. Hypthesis related to the QoL

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 Interface"
"…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."

4.1.2. KSERA QoL Scale

The questionnaires are divide into questionnaires for

- **User acceptance (General impression, Satisfaction, Sustainability)**
- **Physical measurement and ICT based health support abilities**
- **Exercise and Training support abilities**
- **Disease self management abilities**

USERA ACCEPTANCE QUESTIONAIRE

General impression of the KSERA system

Application:

The questionnaire will be used right **after every iteration** of the tests. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

Some questions are listed below. Please circle the number that fits your opinion best.

- **What is your general impression of the KSERA system?**
- **Was it difficult to learn how to use and interacte with the KSERA system?**

yes 1 2 3 4 5 not at all

- **Was it difficult for you to use and interact with the system without help?**

yes 1 2 3 4 5 not at all

- **Were the instructions given by the system (NAO) clear, understandable and easy to follow?**

yes 1 2 3 4 5 not at all

- **Do you think the KSERA System is able to help you in case of an emergency?**

yes 1 2 3 4 5 not at all

- **Did you feel being observed by the system and especially NAO?**

yes	1	2	3	4	5	not at all

Evaluation of the feeling right after the test

Application:

The questionnaire will be used right **after every iteration** of the tests. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the emotions regarding the way you are **feeling right now** about the KSERA SYSTEM and about the general feeling about such a system.*

*Under each question there is a scale describing to what degree you may/may not feel each emotion. **Please circle the number corresponding to the statement that best describes how you are feeling RIGHT NOW about the system!***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

8. It is fun.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

9. It is unpleasant.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

10. I dislike it.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

11. I am feeling cheerful.

1	2	3	4	5

Not at all	A little	Almost	Quite	Very much

12. I am feeling tired.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

13. I am feeling refreshed.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

14. I am feeling stressed.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

15. I am feeling calm.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

16. I am feeling bored.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

Satisfaction Evaluation

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

Some statements are listed below. Please circle the number that fits your opinion best.

1. How beneficial do you believe is KSERA for you?

1	2	3	4	5
Not at all beneficial	Not very beneficial	Neither beneficial nor harming	Beneficial	Very beneficial

2. How often would you use KSERA if you had it at home?

1	2	3	4	5
Never	Less than once a month	A couple of times a month	Once a week	More than once a week

3. Did KSERA make you feel more confident about using new technologies?

1	2	3	4	5
Not at all confident	A little confident	Neutral	Confident	Very confident

4. KSERA was amusing and I enjoyed interacting with it.

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

5. I prefer spending my time by doing something else than interacting with KSERA

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

6. Using KSERA was boring and did not interest me.

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

7. KSERA has met my expectations.

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

Physical measurement abilities

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the way you are **feeling right now** about the possibilities for support of physical measurement performance you have been doing and about the general feeling about such a supported physical parameter measurement**.***

*Under each question there is a scale describing how you think about the different questions. **Please circle the number corresponding to the statement that best describes how you are thinking about the different questions***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

1. Reminding me of the regular measurement would make me perform the physical measurement more often

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

2. By the direct KSERA feedback I would have a better understanding of my physical status compared to the traditional representation on the device

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

3. I would prefer a technical system giving me hints, when to measure my parameters to a measurement timetable or a traditional remindersystem

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

4. It would be a benefit if the KSERA system would log my physical measurement data and give me feedback about the trend

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

5. NAO is motivating me to perform my measurements

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

6. It would be more motivating to perform the measurements if the robot would criticize me.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

7. **It would be annoying for me, if the robot would give me hints all the time.**

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

Exercise and training support abilities

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the way you are **feeling right now** about the physical training you have been doing and about the general feeling about such a supported training**.***

*Under each question there is a scale describing how you think about the question. **Please circle the number corresponding to the statement that best describes how you are thinking about the question***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

1. NAO is motivating me to do the excercises.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

2. I am more motivated, compared to only having a training plan

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

3. I am more motivated, compared to only having a personal trainer

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

4. The excercises were described and shown in an understandable way

1	2	3	4	5

Not at all	A little	Almost	Quite	Very much

5. Mimicing the excercises makes me performing them in a better way, than only having the raw description

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

6. The direct feedback after the exercise performance makes me feel good

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

Disease management abilities

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the way you are **feeling right now** about the physical training you have been doing and about the general feeling about such a supported training**.***

*Under each question there is a scale describing how you think about the question. **Please circle the number corresponding to the statement that best describes how you are thinking about the question***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion

1. KSERA offers the opportunity to get a better overview of possible problem in my disease management

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

2. KSERA is a good support in remembering times and dates of therapies and treatment

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

3. KSERA is a good support in remembering times and dates of therapies and treatment

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

4. KSERA and especially NAO could be a good help in case of any emergency and could help me get help faster

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

5. By regular information support by the KSERA system I would have a better overview of my phasical status.

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

Independent Living >>> only for PT2

1. KSERA makes me feel more autonomous

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

2. KSERA makes me feel I can control my health better

1	2	3	4	5
Strongly dissagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

3.Would you believe that you can use the KSERA system at home without help?

YES NO

For end evaluation only

Did you feel safer with the system?

1	2	3	4	5
Not at all	A little	Almost	Quite	Very much

Did the system disturb your daily routine?

1	2	3	4	5

5. Human Robot Interaction – Perception of the SAR system

KSERA HRI Scale and Godspeed Scale will be used for evaluation special user experience questions concerning the human robot interaction.

5.1. KSERA HRI Scale

The KSERA HRI Scale is a score based scale. Pairs of robot attribute, attitude, behavior and interaction opposites such as, friendly - unfriendly; humanlike - machinelike are presented. The trial Participants rate their expectations of the robot on a 5 point scale. The items can be added to create an overall scale or also as single factors.

Lower scores are associated with more **favourable attitudes** towards the robot)

Domains of HRI treated with the following questionaire

1. **Robot attribute**
2. **Robot attitude**
3. **Robot behavior**
4. **Robot interaction**

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the way you are **feeling right now** about the physical training you have been doing and about the general feeling about such a supported training.*

*Under each question there is a scale describing how you think about the question. **Please circle the number corresponding to the statement that best describes how you are thinking about the question***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

1. Robot attribute

"How would you describe NAO's movements?"

fast	1	2	3	4	5	slow
elegant	1	2	3	4	5	clumsy

humanlike	1	2	3	4	5	machinelike

"How would you describe NAO's speech output?"

understandable	1	2	3	4	5	not understandable
loud enough	1	2	3	4	5	too quiet
humanlike	1	2	3	4	5	machinelike

"How would you describe NAO's speech recognition?"

understandable	1	2	3	4	5	not understandable
loud enough	1	2	3	4	5	too quiet
humanlike	1	2	3	4	5	machinelike

"How old is NAO?"

Infant Toddler Child Teenager Adult

"What gender is NAO?"

female male neither

2. Robot attitude:

questions dealing with **attitude**:

"How well did NAO pay attention to you?"

attentive	1	2	3	4	5	not attentive

"Was NAO friendly?"

friendly	1	2	3	4	5	unfriendly

"Did NAO seem happy?"

happy	1	2	3	4	5	unhappy

"Does NAO seem to be intelligent?"

intelligent	1	2	3	4	5	not intelligent

"Did NAO enter your *"circle of privacy"*?"

no	1	2	3	4	5	too much

"Did NAO enter your *"circle of privacy"*?"

no	1	2	3	4	5	too much

"Would you trust NAO"

yes	1	2	3	4	5	not at all

3. Robot behaviour:

Questions dealing with NAO's **behavior** will include

"Were NAO's actions natural?"

natural	1	2	3	4	5	unnatural

"Did NAO react appropriately to your actions?"

appropriate	1	2	3	4	5	not appropriate

"Does NAO seem alive?" (compare also "godspeed")

alive	1	2	3	4	5	dead

4. Comparison between smart home solutions and KSERA – this is a very KSERA specific evaluation and may not be used without modification for related projects.

Testscenario: Attention Raising

1. **"What system seems to be more motivating for you to really perform the wished/asked task?"**

   KSERA (NAO) LUI (Touch Screen Interface)

2. **"What seems to be preferable for you in terms of steering/choosing functions?"**

   KSERA (NAO) speech/word input LUI (Touch Screen Interface)

3. **"What system seems to be attrtact your attention in a better way?"**

KSERA (NAO) LUI (Touch Screen Interface)

4. **“Why did this system attract your attention in a better way?”**

   KSERA (NAO) speech/word input LUI (Touch Screen Interface)

Testscenario: Exercise Performance with NAO / BEAMER / LLM

1. **“What system for explaining exercises would you prefer?”**

   KSERA (NAO) KSERA (NAO with Beamer) LUI (Touch Screen Interface)

2. **“Why would you prefer this?”**

3. **“What display seems to preferable for you?”**

   KSERA (NAO with Beamer) LUI (Touch Screen Interface)

4. **“What system seemed to motivate you the most?”**

   KSERA (NAO) KSERA (NAO with Beamer) LUI (Touch Screen Interface)

5. **“If you had the coice of taking one system to your home for one month – which one yould you choose?”**

   KSERA (NAO) KSERA (NAO with Beamer) LUI (Touch Screen Interface)

Testscenario: Exercise Performance with NAO / BEAMER / LLM

1. **“What system seems to be preferable for getting information faster?”**

   KSERA (NAO) LUI (Touch Screen Interface)

2. **"What system seems to be preferable for getting better/deeper information?"**

KSERA (NAO) LUI (Touch Screen Interface)

5.2. HRI – Godspeed Questionaire

„The series of questionaires proposes a measurement tool to evaluate the the users' perception of robots. This series shall be called "Godspeed" because it is intended to help creators of robots on their development journey. It is a commonly used measurement instrument for the anthropomorphiscm, animacy, likeability, perceived intelligence and perceived safety of robots"

This questionnaire has been discussed in detailed in [Bartneck2009].

Application:

The questionnaire will be used **once after the first iteration of the trials and once after the last iteration**. The questionnaire will be filled out in a separate research room together with a member of the research team. It will be used for direct interviewing the trial participants. The qualitative and quantitative measure are quoted by the members of the research team.

*Below there are some questions listed concerning the way you are **feeling right now** about the robot NAO*

*Under each question there is a scale describing how you think about the question. **Please circle the number corresponding to the statement that best describes how you are thinking about the question***

You do not have to think about your answers much. Remember: There are no right or wrong answers. Just your opinion.

Godspeed I: Anthropomorphism

Please rate your impression of the robot on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly

Godspeed II: Animacy

Please rate your impression of the robot on these scales:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

Godspeed III: Likeability

Please rate your impression of the robot on these scales:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Godspeed IV: Perceived Intelligence

Please rate your impression of the robot on these scales:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable

Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Godspeed V: Perceived Safety

Please rate your emotional state on these scales:

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised

6. References

[Bartneck2009] Bartneck, C., Croft, E., Kulic, D. & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1) 71-81.

[Davis1989] Davis, F. D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", *MIS Quarterly* 13(3): 319–340

[Murphy2000] Murphy B, Herrman H, Hawthorne G, Pinzone T, Evert H (2000). Australian WHOQoL instruments: User's manual and interpretation guide. Australian WHOQoL Field Study Centre, Melbourne, Australia. World Health Organization (1993). WHOQoL Study Protocol. WHO (MNH7PSF/93.9).

[CCQ] http://www.ccq.nl/ccq.aspx?version=day

[Watson1988] Watson, D., Clark, L., Carey, G., Journal of Abnormal Psychology, Vol 97(3), Aug 1988, 346-353.

[Steinfeld2008] *Steinfeld, A., Common Metrics for Human-Robot Interaction*. Robotics Institute. Carnegie Mellon University. Pittsburgh, PA, 2008.

[ANSI1993] ANSI/AIAA. Guide to Human Performance Measurements,AIAA, Washington, DC, 1993.

[Huang2003] Huang, H., and Messina, E., Toward a generic model for the autonomy levels for unmanned systems. In Proc. NIST Performance Metrics for Intelligent Systems Workshop, (2003).

[Goodrich2003] Goodrich, M., and Olsen, D., Seven principles of efficient human robot interaction. In Proc. IEEE International Conference on Systems, Man and Cybernetics, (2003), 3943-3948.

7. ANNEX – Analysis sheets

In the following section the analysis sheets for the questionaires of the

- KSERA QoL scale
- KSERA HRIO scale
- Godspeed Scale

used during the evaluation phases in the KSERA project are presented.

Yellow marked questions were filled in this sheets taking the results directly from the questionnaires. Red ones were inserted using the inverted result fro the questionnaires. Green sections mark cluster scores.

In addition also an Reliability Evaluation form is presented as an example how the technical evaluation results (in this case for one testcase 1) were analyzed in order to achieve comparable and meaningful results.

KSERA QoL Scale

Description | red questions will be INVERTED, yellow ones will be normal | Different clustersn

				user nr. 1	.....	user nr. i	Resulting
General impression of the system (maximum 5)	corresponding question	factor	cluster score				
	1	general impression					
	2	handling					
	3	ease of use					
	4	understandability					
	5	support ability					
	6	privacy					
			SCORE General impression of the KSERA system - - the higher the better				
				user nr. 1	.....	user nr. i	
Emotianal feeling (maximum 7)	corresponding question	factor	cluster score				
	1	fun					
	2	unpleasentness					
	3	dislike					
	4	happyness					
			SCORE Likeability of KSERA - How does the person feel about KSERA - corresponding to how the person likes KSERA - the higher the better				
	1,2,3,4						
	5	feeling strong					
	6	feeling tired					
	7	feeling fresh					
	8	feeling stressed					
	9	feel calm					
	10	boored					
			SCORE for feeling after the test - the higher the better				
	5,6,7,8,9,10						
				user nr. 1	.....	user nr. i	
Level of satisfaction (maximum 5)	corresponding question	factor	cluster score				
	1	usable					
	2	frequency of use					
	3	use of new technology					
	4	entertainment					
	5	system denial					
	6	system boorness					
	7	expectancy					
			SCORE for substentativeness of KSERA				
	4.6		SCORE for periodicity of KSERA use				
	2,5,6						
	1.2		SCORE for closure of digital divide				
Physical parameter measurement abilities (maximum 5)	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i	
	1	user need for frequent monitoring					
	2	KSERA perceivability score (compared to actual monitoring tools use)					
	3	KSERA user need score (compared to actual monitoring tool)					
	4	user need for trend calculation					
	5	KSERA motivation score					
	6	KSERA motivation score (with additional criticism)					
	7	user need for frequent monitoring					
			SCORE concerningNAO motivation/criticism				
	6.7		SCORE for the user need for monitoring support				
	1,3,7		SCORE concerning advantage of NAO for physical parameter measurement				
	3.4		SCORE for disease management abilities				
	2,4,5						
Trainig support (maximum 5)	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i	
	1	KSERA motivation					
	2	KSERA motivation score (compared to trainingsplan)					
	3	KSERA motivation score (compared to human trainer)					
	4	excercise perceivability					
	5	KSERA perceivability score (compared to only description)					
			SCORE for training support abilities compared with SOTA				
	2,3,5		SCORE for motivated and easy to use training support abilities of KSERA				
	1.4		SCORE for training support				
	1,2,3,4,5						

©KSERA consortium

KSERA HRI Scale

Description		red questions will be INVERTED, yellow ones will be normal	Different clustersn					
characteristics of the robot	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i	**SCORE median**	**SCORE median normalized**
	4a	speed						
	4b	elegance						
	4c	humanoid - MOVEMENT						
	5a	understandability						
	5b	loudness						
	5c	humanoid - SOUND						
	6	speech recognition						
	4c, 5c		SCORE for humanoid perception (max 10)					
	4a, 4b, 4c		SCORE for movement perception (max 15)					
	5a, 5b, 5c		SCORE for conversational abilities perception (max 15)					
Attitude towards the robot	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1	mindfulness						
	2	friendlyness						
	3	happiness						
	4	intelligence						
	5	distance						
	6	confidence						
	1,2,3,6		SCORE politness perception (max 20)					
	1,2,3,4,5,6		SCORE for attitude towards the robot (max 30)					
behaviour of the robot	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1	naturalness						
	2	adequateness						
	3	animated						
	1,2,3		SCORE for robot attitude (max 15)					
Age of NAO	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
		age						

Godspeed Scale

Description		red questions will be INVERTED, yellow ones will be normal	Different clustern					
Godspeed I:Anthropomorphism	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i	**DIA per question**	**SCORE median normalized (maximum = 5)**
	1 2 3 4 5							
	1,2,3,4,5		**Score for Antropomophism**					
Godspeed II: Belebtheit	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1 2 3 4 5 6							
	1,2,3,4,5,6		**Score for Belebtheit**					
Godspeed III: Sympathy	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1 2 3 4 5							
	1,2,3,4,5		**Score for Sympathy**					
Godspeed IV: perceived intelligence	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1 2 3 4 5							
	1,2,3,4,5		Score for perceived intelligence					
Godspeed V:feeling of individual safety	corresponding question	factor	cluster score	user nr. 1	.....	user nr. i		
	1 2 3							
	1,2,3,		Score for individual safety					

Reliability Evaluation Form - Testcase 1

		TP1 TR1	TR2	TR3	... TR1	TR2	TR3	TP I TR1	TR2	TR3	%	min	mean	max	Sum	Anzahl criteria
Testcase 1 Injection Point 0 start in tester interface																
Testcase 1																
STC 1 Nao navigates to TS / designated place																
A1	Is Nao performing movements in the correct way?	1			1	1		1	0		50%				4	5 yes if no turning necessary
A2	How long does it take NAO to perform the intended movement?	20			45	28		31	33			20	31	45		5 from getting up until search for face starts
A3	Is the KSERA system calculating the right user position?	1			1	1		1	0		50%				4	5 yes if square marker stays at least partially over the user during the time it takes nao to nav
A4	Is the KSERA system calculating the right NAO position?	0			1	1		1	1		50%				4	5 yes if the marker does not loose nao during the walk (direction of movement not considered)
																0
																0
Testcase 1 Injection Point 1 start in tester interface in case STC1 failed																0
																0
STC 2 Nao detects focus of the TS																0
A1	Is Nao performing movements in the correct way?	1			1	1		1	1		100%				5	5 yes if head movements correct (no matter which direction it turns first and how fast)
A5	Is KSERA calculating the right user's gaze direction?															0
A14	Is KSERA able to detect the users face	1			1	1		0	0		50%				3	5 yes if the system was able to find the user at least during face search. No if not (too far awa
A2	How long did it take until the system found the users face?	30			26	46		5	90			5	39	90		5 from beginning the search (head movement) to face displayed in the gui
																0
In case NAO did not raise the attention successfully (TP was briefed not to look at Nao)																0
																0
STC 3 Nao attracts attention																1
A1	Is Nao performing movements in the correct way?	0			1	1		0	0		40%				2	5
A9	Is Nao "saying" the expected sequence?	1			1	1		1	1		100%				5	5
A10	Are Nao's words understandable by the user?	1			1	0		1	1		50%				4	5 users opinion or testers opinion if not mentioned by the user
A12	Is NAO motivating the user to perform a task?	1			1	1		1	1		100%				5	5 yes if user takes the measurement device
																0
Testcase 1 Injection Point 2 start in tester interface in case STC2 failed																0
																0
STC 4 Nao asks / speaks - (Asks to perfom medical measurement)																0
A9	Is Nao "saying" the expected sequence?	1			1	1		1	1		100%				5	5
A10	Are Nao's words understandable by the user?	1			1	1		1	1		100%				5	5
																0
STC 8 Word recognition																0
A8	Does the KSERA system understand the user command correctly	0			0	0		0	0		0%				0	5 taken from shell output, ok if right decision ("yes" instead of "yes please" is fine)
STC 5 User performs medical measurement																0
																0
B1	Is the KSERA system able to gather medical data				1	1		1	1		100%				4	4
B2	How long does it take until measurement data is received				2	2		2	2			2	2	2		4
Bxx	Is the user able to handle the medical device?				1	1		1	1		100%				4	4
																0
STC 7 Valid Data was received																0
B4	Is the system able to store gathered data, retreive it, act upon it				1	1		1	1		100%				4	4
Testcase 1 Alternative solution TS declines measurement - use these only in alternative scenario																0
																0
STC 4 Nao asks / speaks - (Nao responds to declination of measurement)																0
A9	Is Nao "saying" the expected sequence?															0
A10	Are Nao's words understandable by the user?															0
																0
																0
Testcase 1 Injection Point 3 or 4 (only one of the cases possible depending whether or not valid data was received)																0
																0
STC 4 Nao asks / speaks - (Nao responds - Thank you / Sorry no data)																0
A9	Is Nao "saying" the expected sequence?				1	1		1	1		100%				4	4
A10	Are Nao's words understandable by the user?				1	1		1	1		100%				4	4
																0
Testcase 1 Injection Point 5																0
																0
STC 1 Nao navigates to TS / designated place																0
A1	Is Nao performing movements in the correct way?															0
A2	How long does it take NAO to perform the intended movement?															0
A3	Is the KSERA system calculating the right user position?															0
A4	Is the KSERA system calculating the right NAO position?															0
	General Questions after the Testcase															0
C2	Has the KSERA system performed timing operations correctly?	1			1	0		0	0		40%				2	5
C3	Is KSERA able to perform the correct sequence of tasks in the right order?	0			1	0		0	0		20%				1	5
C4	Is KSERA able to derive the correct decisions based on input data?	0			1	1		1	1		80%				4	5
xx	Ev. Did you have eye contact with NAO?	1			1	1		1	1		100%				5	5
xx	**Has the KSERA System performed the Testcase correctly?**	0			0	0		0	0		0%				0	5
																120