in: Proc. Int. Conf. on Social Robotics (ICSR), LNAI 9979, pp. 890-99, 2016

Contribution towards Evaluating the Practicability of Socially Assistive Robots – by Example of a Mobile Walking Coach Robot

Horst-Michael Gross, Markus Eisenbach, Andrea Scheidig, Thanh Quang Trinh, and Tim Wengefeld

Ilmenau University of Technology, Neuroinformatics and Cognitive Robotics Lab
98693 Ilmenau, Germany
horst-michael.gross@tu-ilmenau.de

Abstract. This paper wants to make a further contribution towards a more transparent and systematic technical evaluation of implemented services and underlying HRI and navigation functionalities of socially assistive robots for public or domestic applications. Based on a set of selected issues, our mobile walking coach robot developed in the recently finished research project ROREAS (Robotic Rehabilitation Assistant for Walking and Orientation Training of Stroke Patients) was evaluated in three-stage function and user tests, in order to demonstrate the strengths and weaknesses of the developed assistive solution regarding the achieved autonomy and practicability for clinical use from technical point of view.

Keywords: Socially assistive robotics, Evaluation of practicability

1 Motivation

In the assistive robotics community, more and more researchers are already aware of the challenges involved in studying autonomously behaving interactive systems "in the wild" and follow best practices in studying these robots in natural interaction settings as suggested by [1] and [2]. However, often the setup and the actual implementation of the tests are still not described in a sufficient level of detail, leaving room for speculations, particularly with respect to the achieved autonomy and the practicability of the developed solution from technical point of view. Therefore, the main objective of this paper is to make a further contribution towards a systematic and more transparent technical evaluation of implemented services and underlying basic functionalities of socially assistive robots [3] for public or domestic applications. This way, a more self-critical and honest survey of the strengths and still existing weaknesses of the robot's practicability should be made possible. In the scope of our previous projects, the *ShopBot* TOOMAS [4], the companion robot in CompanionAble [5], the health assistant in SERROGA [6], and the walking coach in ROREAS [7] [8], we gained many experiences in making assistive robotics suitable for real-world applications. Based on these experiences, we have compiled a set of questions that should be clarified in publications dealing with the autonomy and practicability of a developed robotic solution. These issues are divided into the following topics:

Topic S – Spectrum of available services/applications and skills:

S1: What *services/applications* for the users are already available a) working completely autonomously on the robot b) only usable with external sensors (e.g. cameras), c) requiring remote control by a tele-present operator?
S2: What *skills* for navigation and HRI are available for the robot at which level of autonomy?
S3: What kind of IT-infrastructure is required for that on-site? What are the consequences for the later practical application?

Topic M – Maturity level:

M1: What is the *maturity level* of the robot system to be tested? Is it still a demonstrator, a lab prototype, or already a product available on the market?
M2: Have there already been *function tests* outside the lab in the field (when, where, how often, how long, what conditions)?
M3: Have there already been *user tests* with the end users in the final operational environment (when, where, how long, what conditions)?
M4: Was *accompanying personnel* from technical or social sciences staff present during the user tests, and where was the staff while the tests were running?
M5: How long was the robot available for the user, how was the usage rate?

Topic F – Function tests of basic functionalities: Here, the scenario-specific functionalities in navigation and HRI are to be quantitatively evaluated.

F1: What *navigation functionalities* have been tested under what conditions?
F2: Are there *navigation problems* encountered during the tests, and how were these quantified (e.g. number of collisions, deadlocks, close encounters with obstacles, violations of personal space, localization failures, etc.)?
F3: What *HRI functionalities* have been tested under what conditions?
F4: Have there been *HRI-malfunctions* (e.g. in person detection, person tracking, user re-identification, etc.)?
F5: What *success rates* of basic functionalities have been determined (e.g. localization accuracy, target achievements, user detection/search, etc.)?
F6: Were *manual interventions* necessary before and during the tests (e.g. labeling no-go areas, preparing critical obstacles, triggering emergency stops, changes in the application procedure while testing)?
F7: Was the *complexity* of the test environment quantitatively evaluated (e.g. by total floor area, free space, navigable area, clearance, shape factor, mean passage width) to allow for a comparison of the test results? (see [6])

Topic U – User tests at technical application level: Here the level of autonomy, the practicability of the application, and the interplay of the basic functionalities are to be evaluated by the following error measures:

U1: *Uncritical failures*: can be handled by the application itself (e.g. driving to a meeting point if user contact is lost, autonomously terminating a bumper-stop if knowledge about the triggering event is available)
U2: *Critical failures*: can be resolved by remote intervention (see Sec. 3) through an operator (e.g. correction of a wrong person re-identification hypothesis)
U3: *Very critical failures*: cannot be resolved by remote intervention through an operator (e.g. sensor failures or deadlocks after collisions).

In the following sections, when describing our test strategy and the achieved results and observed problems, we make use of links to these issues, e.g. (↗ F3).

2 Mobile Walking Coach Robot

Based on this set of issues, the recently finished research project ROREAS (Robotic Rehabilitation Assistant for Walking and Orientation Training of Stroke patients) [7], [8] was systematically evaluated. The ROREAS project aimed at developing a robotic rehabilitation assistant for walking self-training of stroke patients in late stages of the clinical post-stroke rehabilitation. Such a walking coach robot is supposed to motivate and accompany stroke patients who already got the permission to walk on their own without professional assistance during their walking exercises in a clinical rehab center (Fig. 1). A specific characteristic of ROREAS is its strongly human-aware, polite and attentive social navigation and interaction behavior [9] as it is necessary for a rehab assistant that can motivate patients to start, continue, and regularly repeat their self-training with joy. In [8], we already described the specifics and challenges of the clinical setting and the technical requirements for the robotic walking coach, presented the ROREAS prototype (Fig. 1), an application-tailored mobile robot developed within the ROREAS project to meet the requirements to a personal training robot (↗ M1), and gave an overview of the robot's system architecture.

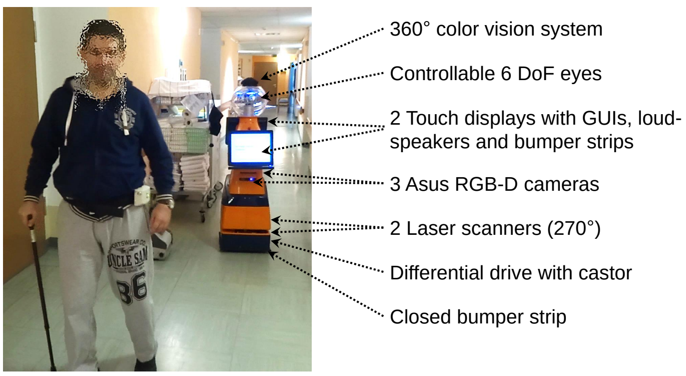

Fig. 1: Robotic walking coach "RINGO" (↗ M1) with its main equipment for environmental perception, navigation, and HRI during a walking tour in our test site, the "m&i Fachklinik" rehabilitation center in Bad Liebenstein (Germany).

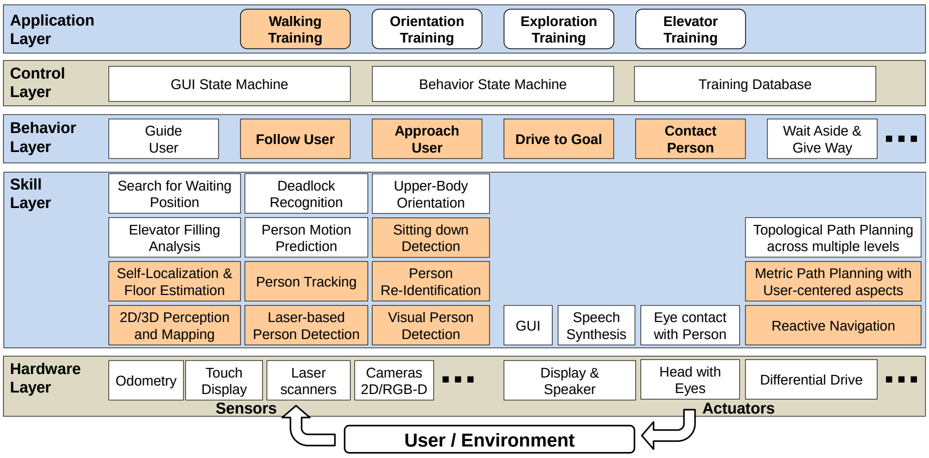

Fig. 2: System architecture of the ROREAS training assistant. Only the reddishly highlighted skills and behaviors are of relevance for this paper, as they are the essential components of the "Walking Training" application, the practicability of which has been evaluated here from technical point of view (↗ S1, S2).

For implementing the training application, numerous robustly functioning basic skills and behaviors were required that had to function completely autonomously in order to achieve the necessary practicability. Therefore, during the function and user tests we particularly paid attention to the skills and behaviors highlighted reddishly in Fig. 2, which have been optimized and evaluated over and over again in order to achieve greater autonomy. In continuation of [8], this paper is focussing on the question how the developed robotic walking coach, its implemented services and the underlying basic functionalities for HRI and navigation can be evaluated systematically to assess its final practicability for the clinical use from technical point of view.

3 Three-stage Approach in Conducting User Tests

Before it was possible to evaluate the walking coach together with stroke patients in user trials, it had to be assured that all the required skills and behaviors for HRI and human-aware navigation (see Fig. 2) did work as expected in the clinical setting. Therefore, we applied a *three-stage approach* in conducting function and user tests with the developed prototype (↗ M1) in the rehab center under everyday conditions. For correction of lacking or wrong decisions of selected skills (e.g. person re-identification [10]) and, with that, for the sake of an interruption-free testing process, we developed a tablet-computer based correction interface connected with the robot by WiFi, which allowed an external test observer to manually correct these decisions from a non-distracting distance (>5 m) (↗M4). By this option for *remote intervention*, the user tests in the clinic could be started earlier than this would have been possible from the readiness level of the respective skills. Moreover, the developers got an objective and situation-specific feedback, in which situation the basic skills and behaviors were still facing problems. Furthermore, this way a direct measure of quality for the autonomous operation of the robot was available, as the number of necessary interventions could be counted (↗ F5, F6). It should be stressed that this option for remote intervention is not to be confused with a robot remote control (which is often used for user studies), as our robot is operating autonomously. The tablet only allows the distant observer to add lacking decisions (e.g. from sitting-down detection), to correct erroneous decisions (e.g. from person re-identification), or to modify the training process in order to keep the training application flowing. An emergency stop can be triggered as well. During the tests, from far distance a second staff member observed problems not detectable by the robot's sensor systems (e.g. violations of personal space, collisions without bumper contact, etc.) and documented them quantitatively. Neither IT-infrastructure of the clinic nor external sensors or markers were required for these tests (↗ S3).

Stage 1 – Functional on-site tests with staff members: To ensure, that all skills and behaviors (see Fig. 2) required by the walking coach do work accurately and securely, first we performed functional on-site field tests with staff members of our robotics lab. These tests were conducted in February 2015 over the course of 4 days and a driven distance of 15,000 meters within several floors of the clinic at different times throughout the day (↗ M2). This was done to assess the robot's basic behaviors under varying conditions, such as challenging building-structures, changes in illumination, and a variable amount of people within the corridors. For quantitative assessment of the skills and behaviors, measures, as e.g. the number of collisions or person mismatches, or the needed travel time, were determined. Regarding the *navigation performance*, the distant test observer counted the number of (i) close (< 10-15 cm) passings of obstacles, (ii) close passings of persons, and (iii) manually triggered emergency stops. Regarding *person recognition during guiding and following*, it was determined, whether and how often the robot confused the current user with a different close-by standing person. A detailed quantitative analysis of the tested skills (↗ F1-F4) and the determined success rates (↗ F5) have already been presented in [8] and are not to be repeated here. In these tests, manual corrections via remote intervention were made only when the re-identification failed or emergency stops were necessary to prevent possible collisions (↗ F6).

Stage 2 – User tests with "patient doubles": After successfully completing the functional tests with staff members of our robotics lab, in May 2015 and shortly before each user test (see below) we evaluated the walking coach again – but this time with the help of clinical staff who imitated the walking behavior of stroke patients (↗ M2, M3). Fig. 3 illustrates the phases of a typical walking training session as it was executed in all following user tests [8]. In these tests, among the stability of the required HRI- and navigation skills (↗ F1, F3) the actual training application and the conclusiveness and comprehensibility of

No.	Phases of a typical walking training session
1.	**Drive of the robot to the patient room** to a non-blocking waiting position near the door
2.	**Initial sitting-down detection at the starting point** in front of the patient room as trigger for making contact with the robot
3.	**Approaching the patient** by the robot for initiating the interaction
4.	**Dialog for selecting the walking tour** and **learning the user model** for user re-identification during training
5.	**Following the patient** to resting or destination points **using user re-identification**
6.	**User sitting-down detection at resting or destination points** as trigger for resting or terminating the training
7.	**Farewell dialog** for presenting the training results (length of the walking tour, duration of training, average walking speed) in front of the patient's room

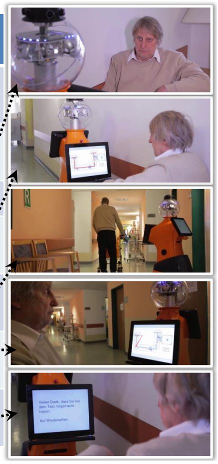

Fig. 3: Phases of a typical walking training session

the training procedure and the necessity of manual remote interventions by the observer ($\nearrow$ F6) were tested. Thus, the level of autonomy and the practicability of the application were in the focus of these trials. A detailed quantitative determination of the success rates of the skills and behaviors ($\nearrow$ F5) and the observed failures at application level ($\nearrow$ U1-U3) was not yet part of these studies, but was left for reasons of manageability for the following user tests.

Stage 3 – Technical user tests with patients: Based on the emulated user tests with staff members and patient doubles, in the period from June 2015 till March 2016 five campaigns of user tests with $N = 26$ stroke patients in total were conducted ($\nearrow$ M3). In all campaigns, only volunteers from the group of stroke patients who already got the permission for doing self-training by their doctor in charge were involved. While the first user trials in June and September 2015 only comprised one predefined short training route, for the following trials in November 2015, January and March 2016 the walking coach was improved to provide freely selectable training routes and, depending on the patients' state of health, sessions with a duration of up to one hour ($\nearrow$ M5). During all user trials, the test observer had the opportunity to correct wrong or lacking decisions in sitting-down detection and user re-identification by remote intervention via the tablet. Our aim was to reduce the number of interventions and to improve the level of autonomy from user tests to user test. Therefore in the subsequent analysis of the user tests, the focus has been directed to the still missing autonomy, in other words, the number of required remote interventions.

4 Results of User Tests with Patients in Stage 3

In the *first user test with patients in June 2015*, all phases of a typical walking training session had to be completed (see Fig. 3), however, it was performed dur-

Criteria	1st user test	5th user test
Period	June 2015 – 2 days	March 2016 – 2 days
Number of patients	5	7
Used walking aids	only walkers	5 x crutch, 2 x walker
Number of sessions	11	**14** (7 on 1st day, 7 on 2nd day)
Driven distance Per session	873 m 80 m	6 650 m 475 m
Total training time Training time per patient	62 min 12 min 24 s	6 hours 15 min 53 min
Number of passers-by	**78** (7 per session, 9 per 100 m)	**679** → 48 per session / 10 per 100 m
Number of remote interventions (for Re-ID)	**19** → 2.2/100 m or 3/10 min.	Offensive: **43** → 1.3/100 m; 2/10 min. Cautious: **19** → 0.6/100 m; 1/10 min.

Fig. 4: Results of the first and last (fifth) technical user tests

ing low traffic times to minimize disturbances by uninvolved passer-by. Essential aspects of this user test are characterized in the tabular overview shown in Fig. 4 ($\nearrow$ M3, M5), while the autonomy already achieved in the different phases of a training session is described in Fig. 5 ($\nearrow$ S1, F6).

During these tests, in three very specific situations the robot collided with unexpected obstacles. The reason for this were rotational movements of the robot by which one of its touch displays softly collided with handrails at the walls. These handrails could not be observed by the robot's laser scanners and 3D-cameras due to their mounting height, and were only detected by the bumper strips at the displays, which triggered an immediate stop. So the robot had to be freed from these situations by manual intervention through the accompanying tests observer ($\nearrow$ U3). These problems did not occur during the preceding function tests and user tests with patient doubles, and came as a surprise therefore. However, the patients did not notice this and continued their training. To handle this problem for the following tests, the respective handrails had to be marked in the navigation map as no-go areas ($\nearrow$ F6). Only twice, the robot violated the personal space of a person ($< 15cm$). In both cases, this behavior was hard to avoid due to the traffic on the corridors (persons suddenly stepping out from rooms or closely passing by). 19 times the test observer had to react by remote invention (3 times in 10 minutes on average) ($\nearrow$ F6) to confirm uncertain or to correct wrong hypotheses of the re-identification module ($\nearrow$ U2). Only two of these cases were false decisions, the others were too uncertain and only required confirmation. In this test, it became apparent that the clothing-based re-identification of patients [10] has a higher degree of difficulty compared to the test with staff members and patient doubles. The cause study for the observed failures showed that the field of view of the panoramic head camera used until then was too limited, and the approaching to sitting patients was still sub-optimal, as the distances for a comfortable handling of the touch screen were

Phases	Achieved Autonomy in 1st user test	Achieved Autonomy in 5th user test
1. Drive to the patient room	100% autonomous	100% autonomous
2. Initial sitting-down detection at starting point	100% by remote interventions	85% autonomous 15% by remote interventions
3. Approaching the patient	100% autonomous	100% autonomous, 47% successful
4. Dialog for tour selection	only 1 training route	free selection from 3 routes
5. Following the patient using user re-identification (ReID)	largely autonomous; ReID: 3 per 10 min. by remote intervention	largely autonomous; ReID: 1 per 10 min. by remote intervention
6. User sitting-down detection at resting /destination points	100% by remote interventions	77% autonomous 23% by remote interventions
7. Farewell dialog	100% autonomous	100% autonomous

Fig. 5: Achieved autonomy in all phases of a training session during the tests

too large. It became apparent, that the training scenario had to be expanded in terms of longer training routes and more options for the patients. Moreover, further remote correction options for the control of the training process should be added to better correct occurring failures immediately during the tests to keep the training flowing. Following this test strategy, in the period between September 2015 and January 2016, three more user tests with $N = 14$ volunteer patients were conducted to assess the improvement of the navigation and HRI skills, that unfortunately cannot be reported here due to lack of space.

In the *last user test in March 2016*, the following issues were in the focus of the practicability investigation: stability of the distance to the accompanied patient, quality of user sitting-down detection and re-identification after installation of a new high resolution panoramic color camera with large vertical field of view (consisting of 6 HD cams), and approaching sitting persons. Of special significance was the question, how the overall behavior of the walking coach and its practicability will be changed, when the remote interventions by the test observer will be more and more restricted to really critical exceptional cases only (very cautious use) in comparison to an offensive use (↗ F6). Essential aspects of this user test are also characterized in the tabular overview in Fig. 4 (↗ M3, M5), while the achieved autonomy is described in Fig. 5, right. (↗ S1, F6).

In the case of an offensive use of interventions (on the first day of the trials) in 43 situations (2 per 10 minutes, or 1.3 per 100 m) remote interventions were carried out (↗ U2) for user re-identification. Reasons for that were missing detections, temporary occlusions of the patients after turning round the corner, or non-detections after taking a seat. In 5 of these 43 cases, there was a confusion with a person closely passing by, and in 2 cases there was no evident reason visible for the hasty intervention of the test observer. By comparison, when a cautious use of the interventions was applied (on the second day), only in

19 situations interventions were necessary (↗ U2), that is 0.6 per 100 m, or once per 10 minutes. In 16 cases, the patients could not be re-identified due to missing detections of the person detector or temporary occlusions by other persons or obstacles, and in 3 cases there were mismatches with passers-by. Collisions with obstacles did no longer occur in this test. 30 times (0.4 per 100 m) the robot violated the personal space of other persons ($< 15cm$), however, without touching them. Again, this was and will be hard to avoid due to the traffic on the corridors, missing person detections (mostly for patients using wheel chairs), and persons suddenly stepping out from rooms or closely passing by (↗ F4).

A robust sitting-down detection is required for the start of the training and for unplanned breaks during the training. From the 14 initial situations (7 per day), 12 were detected autonomously (85%), 2 were not detected due to missing person detections and had to be triggered by remote intervention (↗ U2). The sitting-down detection during the breaks at resting points was necessary in 97 cases, 75 cases were detected autonomously (77%), 22 (23%) still had to be triggered by remote intervention. Reasons for that were temporarily absent person detections because of very atypical views during sitting-down. 111 times the robot had to approach the users while they were sitting. Only in 52 of the final approach positions (47%), the patients could conveniently operate the touch-display. In 59 cases (53%) the approaching had to be terminated by the robot, because the used walking aids blocked the way to the patient, or the robot had problems to robustly detect and track some of the sitting patients.

In all sessions of this test, the average distance between the walking coach and the patient was 1.4 m for the two slowly walking patients and 2.5 m for the five faster walking patients (min. 1 m, max. 4 m) (↗ F5).

5 Conclusions and Outlook

Following our three-stage approach in conducting the function tests and technical user tests in the clinical environment, we have reached a status, in which all components are integrated and most skills and behaviors do function autonomously without any corrections by remote interventions. Nevertheless, there are three important skills and behaviors, the visual person re-identification, the sitting-down detection, and the approach a sitting user behavior, that need to be further advanced by algorithmic improvements to guarantee an autonomous walking training without external support and interventions. So, we have to conclude that the walking coach has not yet reached a maturity level which would allow autonomous operation with patients in the clinical setting (↗ M1). Therefore, instead of only focusing on improving the correctness of all skills, in a subsequent project we follow a more promising strategy to better handle missing or wrong detections, unexpected situations, or still latent shortcomings in the training procedure. So, we are currently implementing a recovery strategy for the most critical case of contact loss to the patient as a result of a failed re-identification. In such a situation, the patient is asked to wait at the next resting point along his tour to give the robot the chance to search for him in this area or on the way to this goal (↗ U1).

Despite these difficulties, for the social science studies that were running in parallel in stage 3 with the $N = 26$ volunteer stroke patients, by the option of doing remote interventions in case of lacking or wrong decisions we could provide a full-value training assistant allowing to evaluate the *usability* and *friendly usage* of the training application by the patients. In this way, without any restrictions we were able to see how well the robot's behaviors and offered training service did fit into the self-training concept. The results of these studies show that the patients and fellow patients were very open-minded and accepted the developed robotic trainer. The robot motivated them for independent training and leaving the room, despite difficulties of orientation, provided a very self-determined training regime, and encouraged them to expand the radius of their training in the clinic. So, a statement frequently repeated by many patients after training with the robot was: "*I have never gone this far alone.*" [11] This makes us optimistic that such a robot coach could bridge the gaps between therapeutically assisted training, independent self-training in the clinic, and training at home.

References

1. Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M.: Common metrics for human-robot interaction. In: Proc. HRI, 3340 (2006)
2. Bonsignorio, F., del Pobil, A. P.: Toward Replicable and Measurable Robotics Research. IEEE Robotics and Automation Magazine. 3: 32-35 (2015)
3. Feil-Seifer, D., Mataric, M.: Socially Assistive Robotics. IEEE Robotics and Automation Magazine, 1:24-31 (2011)
4. Gross, H.-M., Boehme, H.-J. et al.: TOOMAS: Interactive shopping guide robots in everyday use - final implementation and experiences from long-term field trials. In: Proc. IROS, pp. 2005-12 (2009)
5. Gross, H.-M., Schroeter, C. et al.: Further progress towards a home robot companion for people with mild cognitive impairment. In: Proc. SMC, 637-44 (2012)
6. Gross, H.-M., Mueller, S. et al.: Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments. In: Proc. IROS, pp. 5992-99 (2015)
7. Gross, H.-M., Debes, K. et al.: Mobile robotic rehabilitation assistant for walking and orientation training of stroke patients: A report on work in progress. In: Proc. IEEE-SMC, pp. 1880-87 (2014)
8. Gross, H.-M., Scheidig, A. et al.: ROREAS - Robot coach for walking and orientation training in clinical post-stroke rehabilitation: Prototype implementation and evaluation in field trials. Autonomous Robots (2016)
9. Trinh, T. Q., Schroeter, Ch., Gross, H.-M.: "Go ahead, please": Recognition and Resolution of Conflict Situations in Narrow Passages for Polite Mobile Robot Navigation. In: Proc. ICSR, pp. 643-53 (2015)
10. Eisenbach, M., Vorndran, A. et al.: User Recognition for Guiding and Following People with a Mobile Robot in a Clinical Environment. Proc. IROS, 3600-07 (2015)
11. Meyer, S., Fricke, Ch.: Robot Companions for Stroke Therapy - Studying the Acceptance of Assistive Robotics among 80 Patients in Neurological Rehabilitation. In. Proc. Kongress "Zukunft Lebensräume" (in German), pp. 16-24 (2016)