Archives of Gerontology and Geriatrics 87 (2020) 103996

Contents lists available at ScienceDirect

Archives of Gerontology and Geriatrics

journal homepage: www.elsevier.com/locate/archger

Improving gesture-based interaction between an assistive bathing robot and older adults via user training on the gestural commands

Christian Werner[a,b,*], Nikos Kardaris[c], Petros Koutras[c], Athanasia Zlatintsi[c], Petros Maragos[c], Jürgen M. Bauer[a,b], Klaus Hauer[b]

[a] Center for Geriatric Medicine, Heidelberg University, Heidelberg, Germany
[b] Agaplesion Bethanien Hospital Heidelberg, Geriatric Center at the Heidelberg University, Heidelberg, Germany
[c] Institute of Communication and Computer Systems (ICCS), National Technical University of Athens, Athens, Greece

ARTICLE INFO

Keywords:
Aged
Robotics
Activities of daily living
Human-robot interaction
Gestures
Training

ABSTRACT

Background: Gesture-based human-robot interaction (HRI) depends on the technical performance of the robot-integrated gesture recognition system (GRS) and on the gestural performance of the robot user, which has been shown to be rather low in older adults. Training of gestural commands (GCs) might improve the quality of older users' input for gesture-based HRI, which in turn may lead to an overall improved HRI.
Objective: To evaluate the effects of a user training on gesture-based HRI between an assistive bathing robot and potential elderly robot users.
Methods: Twenty-five older adults with bathing disability participated in this quasi-experimental, single-group, pre-/post-test study and underwent a specific user training (10 − 15 min) on GCs for HRI with the assistive bathing robot. Outcomes measured before and after training included participants' gestural performance assessed by a scoring method of an established test of gesture production (TULIA) and sensor-based gestural performance (SGP) scores derived from the GRS-recorded data, and robot's command recognition rate (CRR).
Results: Gestural performance (TULIA = +57.1 ± 56.2 %, SGP scores = +41.1 ± 74.4 %) and CRR (+31.9 ± 51.2 %) significantly improved over training ($p < .001$). Improvements in gestural performance and CRR were highly associated with each other ($r = 0.80–0.81$, $p < .001$). Participants with lower initial gestural performance and higher gerontechnology anxiety benefited most from the training.
Conclusions: Our study highlights that training in gesture-based HRI with an assistive bathing robot is highly beneficial for the quality of older users' GCs, leading to higher CRRs of the robot-integrated GRS, and thus to an overall improved HRI.

1. Introduction

Bathing disability is one of the first limitations in activities of daily living (ADLs) to occur during aging process (Jagger, Arthur, Spiers, & Clarke, 2001; Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963) representing the strongest predictor of subsequent institutionalization in older adults (Fong, Mitchell, & Koh, 2015). Institutionalized and non-institutionalized older adults require personal assistance in bathing more frequently than for other ADLs (Wiener, Hanley, Clark, & Van Nostrand, 1990). The prevalence of bathing disability in community-living older adults increases with age, ranging from 4.6 to 8.6% in those aged ≥65 years (Wiener et al., 1990) to 20.1 % in those aged ≥85 years (Dawson, Hendershot, & Fulton, 1984). An even much higher prevalence has been documented in nursing homes and personal care

Abbreviations: ADLs, activities of daily living; BI, Barthel Index; CRR, command recognition rate; ES, effect size; FES-I, Falls Efficacy Scale-International; GC, gestural command; GDS-15, Geriatric Depression Scale, 15 items; GRS, gesture recognition system; HD, high definition; HRI, human-robot interaction; ICC, intraclass correlation coefficient; MMSE, Mini-Mental State Examination; OGP, observation-based gestural performance; RGB-D, red green blue-depth; SD, standard deviation; SGP, sensor-based gestural performance; SVM, support vector machine; SPPB, Short Physical Performance Battery; STAM, Senior Technology Acceptance Model, T1, pre-test; T2, post-test; TULIA, Test of Upper Limb Apraxia

* Corresponding author at: Agaplesion Bethanien Hospital Heidelberg, Rohrbacher Str. 149, 69126, Heidelberg, Germany.
E-mail addresses: christian.werner@bethanien-heidelberg.de (C. Werner), nkardaris@mail.ntua.gr (N. Kardaris), pkoutras@cs.ntua.gr (P. Koutras), nzlat@cs.ntua.gr (A. Zlatintsi), maragos@cs.ntua.gr (P. Maragos), juergen.bauer@bethanien-heidelberg.de (J.M. Bauer), khauer@bethanien-heidelberg.de (K. Hauer).
@WernerChris84 (C. Werner)

https://doi.org/10.1016/j.archger.2019.103996
Received 18 July 2019; Received in revised form 21 November 2019; Accepted 7 December 2019
Available online 13 December 2019
0167-4943/

facilities, with ≥90 % of residents who require some assistance in bathing (Jones, Dawyer, Bercovitz, & Strahan, 2009; Wiener et al., 1990). As a consequence of the demographic change, the number of older adults in need for bathing assistance will increase, which in turn will contribute to an increase in the burden to both the formal health and social care system and the informal care system. Because bathing is highly sensitive and intimate, it is not unusual for older adults to be reserved against or avoid, as long as possible, personal bathing assistance from caregivers (Ahluwalia, Gill, Baker, & Fried, 2010). In this context, assistive bathing robots that collaboratively support older adults to take care of themselves in bathing can foster independent living, preserve dignity and privacy, and reduce the burden of caregivers.

Human-robot interaction can be defined as "information and action exchanges between human and robot to perform a task by means of a user interface" (International Organization for Standardization, 2012). To enable humans and robots to successfully perform tasks in a collaborative way, an adequate and efficient HRI interface needs to be implemented, making the interaction as natural, intuitive and easy as possible to use, preferably with a minimum of training. There are various ways to communicate and/or interact with a robot (e.g., speech, body posture, gestures, facial expressions, etc.) (Goodrich & Schultz, 2008). Previous studies suggest that older adults tend to appreciate communication methods that resemble natural interactions between humans (Begum, Wang, Huq, & Mihailidis, 2013; Fischinger et al., 2016). Being the most natural and simplest way in human communication, verbal communication is frequently used for HRI interfaces, enabling robots to identify voice commands of the user (Mavridis, 2015). In typical real-world scenarios, voice commands can, however, be disturbed by noise, reverberations, and other interfering sound sources (Alameda-Pineda & Horaud, 2015). Addressing this issue of speech-based HRI and given that gestures also play a central role in human communication (Goldin-Meadow & Alibali, 2013), gesture-based HRI has become a core element in the development of natural, intuitive and easy to use HRI interfaces (Hernandez-Belmonte & Ayala-Ramirez, 2016).

Gestures can be defined as a form of non-verbal communication in which visible bodily actions, typically of the hands and arms, communicate particular messages (Kendon, 2004; McNeill, 1992). For interacting with an assistive robot via gestures, several cognitive abilities are relevant such as attention control, working memory, information processing speed, executive function, and visuospatial abilities. However, most of these abilities show a pronounced decline across the life span into old age (Craik & Salthouse, 2008; Harada, Natelson Love, & Triebel, 2013). In addition, cognitive impairment is frequent among older adults with ADL limitations (Gure, Langa, Fisher, Piette, & Plassman, 2013; Hakkinen et al., 2007), representative of potential end users of assistive robots, which may considerably impede the interaction with such robots, as it has previously been reported also for interacting with other technologies (Schmidt & Wahl, 2019).

Research on gesture-based HRI often seems to focus on improving a robot's technical performance and robustness in recognizing and interpreting a user's input by integrating new hardware evolutions and/or developing new software algorithms (Guler et al., 2016; Liu & Wang, 2018; Wang, Kläser, Schmid, & Liu, 2011). However, successful HRI is not just a matter of the performance of the robot-integrated gesture recognition system (GRS), but also of the quality of a user's input and the characteristics of a user. Thus, to fully understand what makes interaction between humans and robots successful and how HRI can be improved in a broader context, a more in-depth understanding also of the human side of the HRI seems to be necessary. For example, a previous study on gesture-based HRI with assistive mobility robot reported rather poor HRI in frail older adults with some levels of cognitive impairment, with a command recognition rate (CRR) of the robot-integrated GRS of only 40 % (Efthimiou et al., 2016). The low gestural performance observed in a considerable portion of the sample (26 %) has been implicated as one major cause of the low HRI in this study, which therefore called for training approaches on HRI in older robot users to ensure successful HRI.

Training procedures used to teach naïve individuals how to interact with the robot provide a potential option to improve not only the performance of a user's input for HRI but also the user's attitudes and emotions toward the robot (Engelhardt & Edwards, 1992; Louie, McColl, & Nejat, 2014), which have been shown to improve over time of robot use (Stafford, MacDonald, Jayawardena, Wegner, & Broadbent, 2014; Wu et al., 2014) and to be predictive for the quality of HRI (Broadbent et al., 2010). User training on the HRI that takes into account the individual resources and limitations of the user might especially be of importance in older adults, who typically have less technology experience and express more negativity and anxiety toward robot assistance than younger people (Dyck & Smither, 1994; Scopelliti, Giuliani, D'Amico, & Fornara, 2004). The lack of training or advice on how to use new technologies can significantly affect older adults' acceptance of technology (Tacken, Marcellini, Mollenkopf, Ruoppila, & Széman, 2005). For example, the user's perception of his/her own insufficient user performance for HRI associated with a low efficiency in controlling the functionalities of the robot can potentially reduce the self-efficacy and reinforce the feeling of loss of control (Hauer, 2018).

The variability in physical, cognitive, sociological and psychological characteristics increases with age (Hunter, Pereira, & Keenan, 2016; Nelson & Dannefer, 1992; Yang & Lee, 2010). Older adults may thus be regarded as the most heterogeneous population of all. A recent systematic review suggests that previous studies most frequently failed, however, to consider the participant characteristics when studying the interaction of older adults with a robot and highlights the importance for future studies to better examine HRI in later life (Zafrani & Nimrod, 2018).

The primary aim of this study was to evaluate the effects of a specific user training on gesture-based HRI between an assistive bathing robot and potential robot users. We hypothesized that such a training would improve both the gestural performance of the participants and the performance of the robot-integrated GRS, leading to an overall improved HRI. Secondary aims were to explore participant characteristics associated with the initial gestural performance and the training response in the gestural performance, and to examine the relationship between the gestural performance and performance of the robot-integrated GRS. We expected lower cognitive abilities and more negative feelings toward technology to be significantly associated with lower gestural performance (i.e. user input for HRI). According to the rate-dependency phenomenon and general training principles which indicate that intervention response rates are highest in those individuals with the lowest baseline performance (Dews, 1977; Haskell, 1994; Snider, Quisenberry, & Bickel, 2016), we hypothesized that training response in the gestural performance would be significantly associated with the initial gestural performance before training. Moreover, as we assumed that the performance of the robot-integrated GRS in recognizing the user's gestural commands (GCs) would highly depend on the user's gestural performance, we expected better gestural performances to be significantly associated with better performance of the GRS.

2. Methods

2.1. I-SUPPORT bathing robot

The assistive bathing robot used in this study represented a first prototype developed in the I-SUPPORT project (ICT-Supported Bath Robots), which focused on the development of an innovative, modular, information and communication technology (ICT)-supported domestic service robotic system that safely assists frail older or disabled individuals in various bathing tasks (e.g., pouring water, soaping, scrubbing, drying), with the overall aim to promote their independence

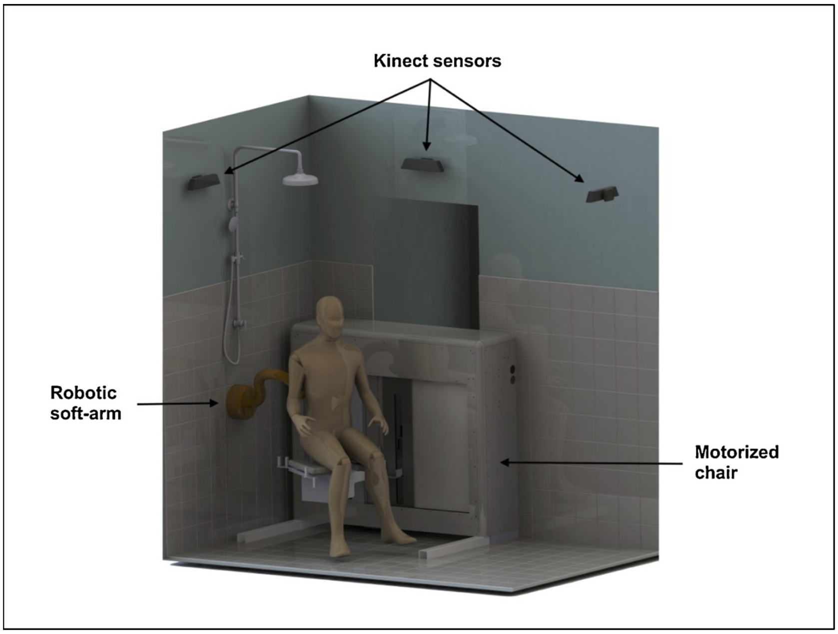

Fig. 1. Rendering of the I-SUPPORT bathing robot placed within the test environment (= typical bathroom of a rehabilitation clinic at a German geriatric hospital).

in this intimate ADL and to relieve the care burden of family caregivers or formal caregivers in medical centers and assisted living environments. More detailed and comprehensive information on the I-SUPPORT project can be found at the project website (http://www.i-support-project.eu/).

The I-SUPPORT bathing robot (Fig. 1) consists of the following main components: (1) a motorized chair for supporting stand-to-sit and sit-to-stand transfers and the transition into and out of the shower area; (2) a robotic soft-arm for the specific bathing tasks (e.g., pouring water, soaping, scrubbing, drying); (3) three Kinect V2 RGB-D sensors and eight condenser microphones for natural audio-gestural HRI (human and robot pose estimation, command and action recognition), and (4) a context-aware system for monitoring environmental (water flow and temperature, air temperature, humidity and illumination sensors) and user information (smartwatch for user identification and (in-)activity tracking) (not provided in Fig. 1). An overview of further technical details of the I-SUPPORT bathing robot will be published elsewhere.

2.2. I-SUPPORT user group

The intended users of the I-SUPPORT bathing robot are persons with (1) dependence in bathing activities and (2) no severe cognitive impairment (Werle & Hauer, 2016). The criteria for dependence in bathing activities was defined according to the bathing item of the Barthel Index (BI) (bathing item: 0 pt. = "patient can use a bath tub, a shower, or take a complete sponge bath only with assistance or supervision from another person") (Mahoney & Barthel, 1965). No severe cognitive impairment was defined as a score of > 17 points on the Mini-Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975).

2.3. Gesture-based human-robot interaction

The I-SUPPORT bathing robot allows the interaction of the users with the robot through a predefined set of GCs for different bathing tasks (e.g., washing, scrubbing). The system architecture for gesture-based HRI consists of three Kinects V2 sensors installed at the walls of the bathroom. The Kinect V2 sensor is equipped with RGB-D and infrared sensors that enable to capture the video (Full HD RGB resolution) and depth information (time-of-flight principle) required for the human and robot pose reconstruction and the identification of the user's GCs. Kinect and other similar sensors are frequently employed for markerless motion tracking and visual recognition in robotics (El-laithy, Huang, & Yeh, 2012; Naeemabadi, Dinesen, Andersen, & Hansen, 2018). Two Kinect V2 sensors were placed inside the shower space for estimating the 3-dimensional pose of the human and robot, and one Kinect V2 sensor was placed outside the shower space for recognizing the GCs performed by the user. The processing methods of the visual information provided by the Kinect sensor for gesture recognition follow state-of-the-art computer vision and machine learning approaches for visual feature extraction and classification. In particular, "dense trajectories" are employed for feature extraction (Wang et al., 2011), an approach frequently used for action and gesture recognition (Baraldi, Paci, Serra, Benini, & Cucchiara, 2014; Yamada, Yoshida, Sumi, Habe, & Mitsugami, 2017) and various other visual recognition problems (Afshar & Salah, 2016; Huang, Zhang, & Li, 2016), especially in cases where the available data for training the algorithms is limited. In brief, this method consist in sampling salient points in the video (e.g., from hand edges, etc.) and tracking them through time, which produces a large number of trajectories. These trajectories are processed to extract the motion boundary histogram (MBH) descriptor in the standard bag-of-features framework (Dalal, Triggs, & Schmid, 2006; Zhang, Marszałek, Lazebnik, & Schmid, 2007), resulting in a high-dimensional numeric representation of the video. Finally, using this representation, each gesture is classified as one of the pre-defined GCs using non-linear support vector machines (SVMs) (Schuldt, Laptev, & Caputo, 2004; (Wang , Kläser , Schmid ,& Liu, 2011). More importantly, SVMs can also provide the probability of each video containing the recognized GC (see 2.9), enabling more in-depth analysis. More technical details on the GRS can be found elsewhere (Kardaris, Rodomagoulakis, Pitsikalis, Arvanitakis, & Maragos, 2016; Rodomagoulakis et al., 2016; Zlatintsi et al., 2018).

2.4. Study design

A quasi-experimental, single-group, pre-/post-test study design was used to analyze the effects of the user training on the gesture-based HRI between the participant and the I-SUPPORT bathing robot. The study was conducted between January 25 and February 8, 2018 with approval of the ethics committee of the Medical Faculty of the Heidelberg University (September 27, 2016; S-382/2016) and in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to study inclusion.

2.5. Study population

Participants were recruited from rehabilitation wards of a geriatric hospital, from nursing homes, and from a hospital-associated geriatric rehabilitation sports club. According to the predefined user group of the I-SUPPORT bathing robot, the following two main inclusion criteria were used to recruit participants: (1) dependence in bathing activities (BI, bathing item = 0 pt.) and (2) no severe cognitive impairment (MMSE score > 17 pt.). Further inclusion criteria were: no severe ADL impairment (BI ≥ 50 pt.); independence in bed-chair transfer (BI, transfer item = 15 pt.); no severe neurological, cardiovascular, metabolic, or psychiatric disorders; residence within 15 km of the study center, and written informed consent.

2.6. Test procedure

The I-SUPPORT bathing robot was installed in a typical bathroom of the rehabilitation clinic at a German geriatric hospital. Seven different GCs for the use case “back region shower process” had to be performed by the participants: (1) wash back; (2) higher temperature; (3) lower temperature; (4) scrub back; (5) repeat; (6) stop, and (7) halt. The correct GCs performed by an expert and used as reference standard can be found in online supplementary videos. During the whole testing procedure, the participants were seated on the motorized chair of the robot. Prior to the pre-test (T1), all participants received a brief introduction on the GCs. For each GC, a test administrator presented a large poster with images displaying the key movement elements of the specific GC and also demonstrated each GC once directly in front of the participant. After this brief introduction, the pre-test was performed with the participant. During the testing phase, the administrator subsequently presented the posters once more for each GC and asked the participant to perform the specific GC shown on the poster. After the participant performed a GC, a short brake was made to give the robot the opportunity to respond on the GC. In case of successful gesture recognition, the robot responded after about 3 s with an appropriate audio response (but did not actually perform the corresponding bathing task) and the next GC was tested. If the robot did not recognize the command correctly in this time interval, the test administrator asked the participant to repeat the GC once more. Independent of the robot response, the test procedure was continued with the next GC after such a second trial. This procedure was followed until all seven GCs were tested. After the pre-test was completed, a more extensive training phase on the GCs was performed by the administrator with the participant (see below). Following this training phase, the test procedure as described for the pre-test was repeated once more (T2 = post-test).

2.7. Intervention

Between the pre- and post-test, a training phase (10 − 15 min) on the specific GCs for the HRI with the I-SUPPORT bathing robot was performed with the participants. For this purpose, specific teaching methods and practice conditions, which have already been demonstrated to be effective for motor learning in older people with cognitive impairment (van Halteren-van Tilborg, Scherder, & Hulstijn, 2007; Werner et al., 2017), were used to facilitate learning of the GCs: mirror technique, combining movements with specific associations, haptic assistance, and high repetitions. The administrator sat directly in front of the participant and demonstrated the GC “like a mirror”, that is, if the participants had to use their right hand for a GC, the administrator demonstrated this GC with the left hand. The participants were encouraged to immediately join the demonstration and to simply mirror the administrator’s movements. During the demonstration, the administrator described the gestures by combining it with specific associations (e.g., “Like you would dip a sponge in a water bucket.” [GC: wash my back]; “Like you would push someone away from you.” [GC: stop]) to facilitate learning and memorizing of the GC. If necessary, also haptic assistance was provided by the administrator to ensure correct movement execution of the GC by the participant. Each GC was trained until the participant was able to perform it once correctly.

2.8. Descriptive measures

Sociodemographic and clinical characteristics including age, gender, living situation (community-dwelling vs. institutionalized), falls in the previous year, and ADL status (BI) were documented from patient charts or by standardized interviews. A trained interviewer assessed cognitive status (MMSE) and psychological status for depression (15-item Geriatric Depression Scale [GDS-15], Gauggel & Birkner, 1999; Sheikh & Yesavage, 1986), fear of falling (Falls Efficacy Scale-International [FES-I], Dias et al., 2006; Hauer et al., 2010), and technology acceptance (Senior Technology Acceptance Model [STAM], Chen & Chan, 2014): subscales for attitude towards technology, perceived usefulness, ease of use, gerontechnology self-efficacy, gerontechnology anxiety, and facilitating conditions). Physical performance was measured by the Short Physical Performance Battery (SPPB, Guralnik et al., 1994).

2.9. Outcome measures

The HRI was evaluated from both the human side, by assessing the participant’s gestural performance, and the robot side, by assessing the performance of the GRS in recognizing the GCs.

Gestural performance was evaluated by (1) scores of a standardized clinical observation measure and (2) sensor-based performance scores derived from the GRS-recorded data.

The clinical observation measure was based on the scoring system of the Test of Upper Limb Apraxia (TULIA), which represents an established test for the comprehensive assessment of gesture production (Vanbellingen et al., 2010). Each GC was rated on a 6-point scale ranging from 0 to 5 points, with higher observation-based gestural performance (OGP) scores indicating better gestural performance. The scoring procedure followed a two-step assessment approach. In a first step, the achievement of the overall movement goal of the GC was evaluated, narrowing the range of the scores to either 0 or 1 points (‘movement goal not achieved’), or 2–5 points (‘movement goal achieved’). The movement goal of a GC was considered to be not achieved if errors occurred that seriously affected the trajectory of the gesture. Trajectories were defined as the spatial orientation of the movement including movement plane relative to the individual’s body, joint coordination, and movement shape. If the movement goal of a GC was achieved, a more detailed error analysis was performed in a second step to yield the final score in the upper scale range (2–5 pt.). The detailed scoring method is presented in Table 1. The first step of this two-step assessment approach was directly performed during the test procedure and was used for deciding whether a second trial was given or not, while the more detailed error analysis was performed after the test procedure using the video recordings of the Kinect V2 sensor. The individual scores per GC were finally averaged over all seven GCs to yield a mean score for the overall observation-based gestural performance (OGP_{total}). Test procedure and scoring were consistently performed by the same person across all participants. Intra-rater reliability for scoring the video recordings of the GCs has been established in a pilot study with 8 participants randomly selected out of the total sample. Excellent intra-rater reliability was found with intraclass correlation coefficients (ICC(2,1), absolute agreement) ranging from 0.82 to 0.95.

A sensor-based gestural performance (SGP) score was calculated for each GC by applying Platt scaling (Platt, 2000) to the output of the SVM classifier of the GRS (see 2.3). This method is implemented by the software libraries used for the GRS (Chang & Lin, 2011) and has been

Table 1
Scoring guide for the observation-based assessment of the gestural performance.

Scores	Description of scoring
5 pt.	The movement goal of the gesture *was achieved*. The gesture was correct and identical to the demonstrated gesture.
4 pt.	The movement goal of the gesture *was achieved*, but errors occurred not affecting the trajectory of the gesture (normal movement plane and spatial location of the hand relative to the body, normal joint coordination and movement shape). Movement was too slow, hesitating, robot-like, and/or sloppy with minor spatial errors such as reduced or excessive amplitudes or unprecise location of the hand relative to the body.
3 pt.	The movement goal of the gesture *was achieved*, but errors occurred subtly affecting the trajectory of the gesture (imprecise movement plane relative to the body, inaccurate joint coordination and movement shape), which were corrected. Additions or omissions of movement components (mainly distal) were present. Brief content errors (substitutions, perseverations, pauses) occurred; however, corrections were made in the ongoing movement.
2 pt.	The movement goal of the gesture *was achieved*, but errors occurred subtly affecting the trajectory of the gesture (imprecise movement plane relative to the body, inaccurate joint coordination and movement shape), which were not corrected. Additions or omissions of (main) movement components (mainly distal) occurred without corrections.
1 pt.	The movement goal of the gesture *was not achieved*. Errors occurred seriously affecting the trajectory of the gesture. The final position was false, major errors in the movement plane, spatial position of the hand relative to the body, joint coordination and movement shape. Overshoot and additional movements (mainly proximal) were present or the gesture was performed with the wrong hand; however, the overall movement pattern of the gesture remained recognizable (1 point). Persisting substitutions (related or unrelated to the gesture) and perseverations occur.
0 pt.	The movement goal of the gesture *was not achieved*. No movement, gesture was totally incorrect or so incomplete that it was not recognizable. Seeking and amorphous movements. No temporal or spatial reference to the requested gesture.

thoroughly shown to provide reliable estimates of class membership probabilities (Caruana, Karampatziakis, & Yessenalina, 2008; Niculescu-Mizil & Caruana, 2005). Each SGP score ranged from 0 to 1 (with higher scores indicating better gestural performance) and quantifies the certainty or degree to which a performed gesture can be classified as the respective GC, according to the GRS. A mean score for the overall sensor-based gestural performance (SGP_{total}) was also calculated by averaging the individual SGP per GC over all seven GCs.

The performance of the robot-integrated GRS was evaluated by its command recognition rate (CRR), defined as the percentage of successfully recognized GCs relative to the seven GCs tested. The test administrator noted the (un-)successful recognition of each command directly during the test procedure.

2.10. Statistical analysis

Descriptive data were presented as frequencies and percentages for categorical variables, and median and interquartile ranges (IQR) and/or mean and standard deviations (SD) for continuous variables. If a participant performed two trials for a GC that both were not successfully recognized by the GRS, the trial with the highest observational-based assessment score was used for the statistical analysis of all outcome measures. In all other cases, the trial with the recognized GC was used. Differences in outcome measures between pre- (T1) and post-test (T2) were analyzed using Wilcoxon signed-rank tests. To quantify the magnitude of pre/post-test changes, effect sizes ($ES = Z/\sqrt{N}$) were calculated and interpreted as small (0.1 to < 0.3), moderate (0.3 to < 0.5), large (0.5 to < 0.7), or very large (≥0.7) (Cohen, 1988; Rosenthal, 1996). Associations between (1) participant characteristics (age, gender, cognitive status [MMSE], physical performance [SPPB], psychological status [GDS-15, FES-I, STAM]) and overall gestural performance (OGP_{total}, SGP_{total}) at T1; (2) system (CRR) and overall gestural performance (OGP_{total}, SGP_{total}); (3) participant characteristics (age, gender, MMSE, SPPB, GDS-15, FES-I, STAM, baseline gestural performance [OGP_{T1}, SGP_{T1}]) and relative changes in overall gestural performance (OGP_{total}, SGP_{total}) over the training phase (T1-T2), and (4) relative changes in the system (CRR) and overall gestural performance (OGP_{total}, SGP_{total}) over the training phase (T1-T2) were analyzed using Pearson's, Spearman rank or point-biserial correlation coefficients (r) as appropriate. Relative changes were calculated as: ((post-test score – pre-test score)/pre-test score) × 100). Correlation coefficients were interpreted as trivial (< 0.1), small (0.1 to < 0.3), moderate (0.3 to < 0.5), high (0.5 to < 0.7), very high (0.7 to < 0.9), extremely high (≥0.9) (Cohen, 1988; Hopkins, Marshall, Batterham, & Hanin, 2009). The sample size was calculated to be n = 25, based on an a priori power analysis for Wilcoxon signed rank tests comparing T1 vs. T2 gestural performance scores (Faul, Erdfelder, Lang, & Buchner, 2007), with a two-sided significance level (α) of 0.05, a statistical power (1-β) of 0.80, and a moderate effect size (Cohen's d_z = 0.6). The expected moderate effect size was derived from findings of previous studies that indicated gross motor skill learning in older adults after one session of semantic instruction and demonstration (Voelcker-Rehage and Willimczik, 2006; Cohen's d_z = 0.7–1.7) and in cognitively impaired older adults after a motor learning exercise program including the same teaching methods and practice conditions as used in the current study (Werner et al., 2017; Cohen's d_z = 0.5–1.1). A two-sided p-value of < 0.05 indicated statistical significance. Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA).

3. Results

3.1. Participant characteristics

The study sample included 25 older people (77.9 ± 7.9 years) who all were dependent in bathing (BI, bathing item = 0 pt.). Thirteen (52 %) participants were recruited from the geriatric rehabilitation sports club, seven (28 %) from nursing homes, and five (20 %) from geriatric rehabilitation wards. The MMSE score averaged 25.6 ± 3.1 points, with about half of the participants (n = 13, 52 %) having some cognitive impairment (MMSE 17–26 pt.). The sample showed a slightly impaired ADL status, with a mean BI score of 81.6 ± 8.6 points (Brefka et al., 2019). The SPPB score averaged 6.1 ± 2.9 points, indicating low physical performance potentially associated with lower frailty status and increased fall risk (Guralnik, Ferrucci, Simonsick, Salive, & Wallace, 1995; Pritchard et al., 2017; Veronese et al., 2014). Most participants (88 %) had no clinically relevant depressive symptoms (GDS-15 > 5 pt.). More than the half reported one or more falls in the previous year (n = 14, 56 %). Fear of falling was low (FES-I = 16–19 pt.) in 4 (16 %), moderate (FES-I = 20–27 pt.) in 9 (36 %), and high (FES-I = 23 – 64 pt.) in 12 participants (48 %) (Delbaere et al., 2010). Almost three out of four participants (n = 18, 72 %) reported concerns about falling while taking a shower or bath (FES-I, bathing item > 1 pt.). Technology acceptance was moderate to high, with mean scores on the different STAM subscales in the upper half of the scoring range (see Table 2).

3.2. Training effects on human-robot interaction

Prior to the training phase, the overall gestural performance was low to moderate, with a median OGP_{total} score of 2.4 points (IQR 1.8–2.9) and a median SGP_{total} score of 0.60 points (IQR 0.47-0.67) (Table 3). Only three participants (12 %) performed at least one GC identical to the demonstrated ones without any movement errors (i.e.,

Table 2
Characteristics of 25 participants.

Variables	
Age, years	77.9 ± 7.9
Sex, females	20 (80.0)
Mini-Mental State Examination, score	25.6 ± 3.1
Geriatric Depression Scale, score	2 [1–3]
Falls Efficacy Scale-International, score	28.8 ± 10.0
Recent history of falls	14 (56.0)
Barthel Index	85.4 ± 11.4
Short Physical Performance Battery, score	6.1 ± 2.9
Technology Acceptance, score[a]	
Attitudes towards technology (max. 20 pt.)	14.6 ± 5.0
Perceived usefulness (max. 30 pt.)	19.9 ± 8.4
Ease of use (max. 20 pt.)	10.8 ± 5.0
Gerontechnology self-efficacy (max. 20 pt.)	12.2 ± 5.2
Gerontechnology anxiety (max. 20 pt.)	12.5 ± 6.1
Facilitating conditions (max. 50 pt.)	30.3 ± 10.5
Living situation	
Community-dwelling	18 (72.0)
Institutionalized	7 (28.0)

Data are presented as mean ± SD, n (%), and median [IQR].
[a] Higher scores indicates better attitudes towards technology, higher perceived usefulness, greater ease of use, higher gerontechnology self-efficacy, lower gerontechnology anxiety, and more facilitating conditions.

OGP score = 5 pt.). The performance of the GRS at pre-test was also only moderate, with a median CRR of 85.7 % (IQR 50.0–85.7).

3.3. Correlational results

Baseline gestural performance was significantly correlated with cognitive status (OGP_{total}: $r = .68$, $p < .001$; SGP_{total}: $r = .68$, $p < .001$) and gerontechnology anxiety (OGP_{total}: $r = .59$, $p = .002$; SGP_{total}: $r = .41$, $p = .041$), such that participants with a higher cognitive performance and less anxiety in gerontechnology showed a higher initial gestural performance at pre-test. High to very high correlation coefficients were observed for these significant correlations, expect for that between SGP_{total} and gerontechnology anxiety (moderate correlation). For all other participant characteristics, there were no significant correlations with the baseline gestural performance ($r = |.01\text{-}.33|$, $p = .144\text{-}.949$) (Table 4).

The pre-/post-test change in the OGP_{total} score was significantly and moderately correlated with gerontechnology anxiety ($r = -.41$, $p = .041$), such that those participants with the highest level of gerontechnology anxiety improved most in the overall gestural performance. The SGP_{total} score tend to confirm this association; however, it just missed the level of significance ($r = -.37$, $p = .069$). In addition, lower baseline gestural performance was significantly and highly correlated with training-induced improvements in the gestural performance (OGP_{total}, SGP_{total}: $r = 0.52\text{-}0.67$, $p < .001\text{-}.008$). For all other

Table 3
Training effects on the gestural performance of the participants and the performance of the robot-integrated gesture recognition system.

Variables	T1	T2	% change	*p*-value	Effect size
Gestural performance					
Observation-based performance scores [pt.]					
Wash back	2.3 ± 1.4	3.0 ± 1.7		.026	0.45
	2.0 [1.0–4.0]	4.0 [2.0–4.0]			
Higher temperature	2.2 ± 1.2	3.4 ± 1.2		.001	0.66
	2.0 [1.0–3.5]	4.0 [2.0–4.0]			
Lower temperature	2.9 ± 1.3	3.6 ± 1.3		.017	0.48
	4.0 [2.0–4.0]	4.0 [3.5–4.0]			
Scrub back	2.1 ± 1.9	3.6 ± 1.6		.002	0.63
	1.0 [0.5–4.0]	4.0 [2.5–5.0]			
Repeat	1.4 ± 1.0	2.3 ± 1.2		.001	0.64
	1.0 [1.0–2.0]	2.0 [2.0–3.0]			
Stop	2.2 ± 1.3	3.2 ± 1.2		.005	0.57
	2.0 [1.0–3.5]	4.0 [2.0–4.0]			
Halt	3.1 ± 1.4	4.2 ± 1.3		.001	0.69
	3.0 [2.0–4.0]	5.0 [4.0–5.0]			
Total performance	2.3 ± 0.8	3.3 ± 0.8	+57.1 ± 56.2	< .001	0.84
	2.4 [1.8–2.9]	3.6 [2.9–3.9]	38.1 [18.9–82.1]		
Sensor-based performance scores [pt.]					
Wash back	0.45 ± 0.33	0.54 ± 0.33		.069	0.36
	0.43 [0.06–0.76]	0.63 [0.21–0.84]			
Higher temperature	0.56 ± 0.33	0.77 ± 0.32		.003	0.59
	0.63 [0.18-0.88]	0.95 [0.72-0.98]			
Lower temperature	0.68 ± 0.36	0.80 ± 0.27		.034	0.42
	0.83 [0.32–0.98]	0.94 [0.60–0.98]			
Scrub back	0.35 ± 0.32	0.69 ± 0.30		< .001	0.79
	0.23 [0.04–0.66]	0.83 [0.58–0.90]			
Repeat	0.40 ± 0.40	0.68 ± 0.41		.022	0.53
	0.24 [0.02–0.91]	0.92 [0.24–0.98]			
Stop	0.66 ± 0.38	0.81 ± 0.31		.010	0.51
	0.86 [0.33–0.98]	0.98 [0.82–1.00]			
Halt	0.67 ± 0.33	0.79 ± 0.29		.211	0.25
	0.85 [0.39–0.96]	0.92 [0.78–0.96]			
Total performance	0.56 ± 0.21	0.73 ± 0.18	+51.1 ± 74.4	< .001	0.79
	0.60 [0.47–0.67]	0.80 [0.60–0.87]	27.6 [13.2–47.0]		
GRS performance					
CRR [%]	70.3 ± 24.0	84.6 ± 21.8	+31.9 ± 51.2	.003	0.59
	85.7 [50.0–85.7]	100 [71.4–100]	16.7 [0–40.0]		

Data are presented as mean ± SD and median [IQR]. *P*-values were given for Wilcoxon signed-rank tests. Effect sizes were calculated as Z/√N. GRS, gesture recognition system; CRR, command recognition rate.

Table 4
Correlations of participant characteristics with pre-test gestural performance (T1) and relative pre-post changes in gestural performance (T1-T2).

Participant characteristics	T1		T1-T2: % change	
	OGP_{total}	SGP_{total}	OGP_{total}	SGP_{total}
Age	.07	−.01	−.17	.01
Sex[a]	.21	.26	−.18	−.22
Mini-Mental State Examination	.68***	.68***	−.28	−.24
Geriatric Depression Scale	−.29	−.26	−.02	.05
Falls Efficacy Scale-International	−.06	−.04	−.02	−.01
Short Physical Performance Battery	−.17	−.33	.19	.33
Technology Acceptance				
Attitudes towards technology	.22	.15	.09	.08
Perceived usefulness	.28	.32	.04	.06
Ease of use	.07	.18	−.09	−.16
Gerontechnology self-efficacy	.16	.24	−.13	−.19
Gerontechnology anxiety	.59**	.41*	−.41*	-.37+
Facilitating conditions	.23	.35+	−.03	−.09
Baseline gestural performance				
OGP_{total}			−.67***	
SGP_{total}				−.52**

Correlations were given as Pearson's, Spearman rank or point-biserial correlation coefficients (r) as appropriate.
+ $p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

participant characteristics, there were no significant correlations with the changes in the overall gestural performance ($r = \ | < .01\text{-}.33|$, $p = .109\text{-}.987$).

Very high to extremely high significant correlations were obtained between the gestural performance and CRR at pre-test (OGP_{total}: $r = .94$, $p < .001$; SGP_{total}: $r = .83$, $p < .001$) and post-test (OGP_{total}, SGP_{total}: $r = 0.81$, $p < .001$) (Table 5). The improvement in the overall gestural performance of the participants was also significantly and highly correlated with the improvement in the CRR of the robot-integrated GRS (OGP_{total}, SGP_{total}: $r = 0.80\text{-}0.81$, $p < .001$).

4. Discussion

The present study aimed to provide a more in-depth understanding of the human side of the gesture-based HRI between an assistive bathing robot and potential end-users. Being representative of the potential user group of the bathing robot, we recruited older people with bathing disability and evaluated the effects on the HRI of a user training specifically designed and tailored to the needs and requirements of this population to improve their performance in interacting with the robot using GCs. In addition, we investigated whether the gestural performance and training response were associated with individual differences in participant characteristics and whether training-induced improvements in the gestural performance would lead to a better performance of the GRS in recognizing the participant's GCs.

Table 5
Correlations between the performance of the robot-integrated gesture recognition system and the gestural performance of the participants.

CRR	OGP_{total}			SGP_{total}		
	T1	T2	T1-T2: % change	T1	T2	T1-T2: % change
T1	.94***			.83***		
T2		.81***			.81***	
T1-T2: % change			.81***			.80***

Correlations were given as Pearson's or Spearman rank correlation coefficients (r) as appropriate.
*** $p < 0.001$. CRR, command recognition rate; OGP, observation-based gestural performance; SGP, sensor-based gestural performance.

Our results clearly indicate that the user training was highly beneficial for improving the gesture-based HRI between the assistive bathing robot and the participants. Lower cognitive performance and higher gerontechnology anxiety were identified to negatively affect the participants' initial gestural performance. However, lower cognitive performance did not influence their training response, and higher initial gerontechnology anxiety was associated with even greater benefits in the gestural performance over the training phase. The participants who benefited the most from the user training were those with the lowest initial gestural performance at baseline. For both testing sessions, as well as for the changes between pre- and post-test, the performance of the robot-integrated GRS was found to be closely related to the gestural performance of the participants.

4.1. Training effects on human-robot interaction

Due to the common lack of experience of older adults in interacting with a robot (Smarr et al., 2012, 2014), potential age-related limitations in cognitive abilities relevant for gesture-based HRI, and previous findings on gesture-based HRI for an assistive mobility robot in a similar population (Efthimiou et al., 2016), the initial gestural performance of the participants at pre-test was expected to be rather low. Our results confirmed this expectation, with only low to moderate gestural performance scores and a very small number of participants (3 out of 25) performing any GC without errors after the brief introduction before pre-test. This finding indicates that a single, brief introduction in gesture-based HRI with an assistive robot does not seem to be sufficient to ensure adequate quality of a user's input for such interaction in frail older adults with some levels of cognitive impairment. As the robot-integrated GRS depend on an adequate quality of the user's input, the low to moderate gestural performance was directly translated into an only moderate CRR, leading to an overall rather unsatisfying HRI at pre-test. To overcome these user-related issues of the HRI and to improve the gestural performance of the participants, a user training on HRI was implemented including teaching methods that have already been demonstrated to be effective for learning motor tasks in older people with cognitive impairment (van Halteren-van Tilborg et al., 2007; Werner et al., 2017). Significant improvements in almost all outcome measures – with predominantly large effect sizes – confirmed our primary study hypothesis that such training improves the participants' movement execution of the GCs, leading to an improved CRR of the robot-integrated GRS, and thus also to an overall improved HRI.

Improvements in the gestural performance were documented by different outcome measures. We developed and used a standardized clinical observation measure for which the scoring method was derived from an established and valid clinical test for gesture production (TULIA) (Vanbellingen et al., 2010), as well as sensor-based performance scores recorded by the robot-integrated GRS. The latter was chosen to substantiate the training effects documented by observation-based outcomes by technically measured, more objective outcomes. Further, this approach of using the already existing data flow of the robot-integrated sensing technique for assessment purposes has been recommended for the evaluation of HRI with assistive robots, allowing for highly specific assessments exactly tuned to the robot's functionality to be evaluated (Werner, Ullrich, Geravand, Peer, & Hauer, 2016).

4.2. Correlational results

Consistent with our hypothesis, lower initial gestural performance was significantly associated with lower cognitive status and more negative feelings toward technology, highlighting the relevance of the user's cognitive abilities for gesture-based HRI as well as the previously reported relationship between a user's emotions toward the robot and the quality of HRI (Broadbent et al., 2010). According to that, training programs to improve the quality of a user's input for HRI seem to be of particular importance in older adults with lower cognitive status and higher technology anxiety.

Higher gerontechnology anxiety was identified to be significantly associated with higher training gains in the gestural performance. This might be related to the fact that participants with a general higher anxiety towards technology may have been initially also more anxious toward the assistive bathing robot; however, as emotions toward a robot have been shown to improve with increasing user experience (Engelhardt & Edwards, 1992; Louie et al., 2014), the experience with the robot at pre-test may have reduced the anxiety in these participants, which in turn may have had a beneficial side effect on the gestural performance in participants with higher technology anxiety in addition to the specific training effect.

As hypothesized, the lower initial gestural performance was also a significant factor for higher training gains in the gestural performance, which is in accordance with the rate-dependency phenomenon and general training principles (Dews, 1977; Haskell, 1994; Snider et al., 2016). This suggests that the participants with the lowest initial gestural performance could also be successfully trained in the GCs and even represented those that benefitted most from the user training.

In contrast, improvements in the gestural performance over the training phase were not significantly associated with the participants' cognitive status in our study, suggesting that a positive training response can also be achieved in older adults with mild-to-moderate cognitive impairment. This might especially be explained by the fact that we applied specific teaching methods and practice conditions in the user training which have been shown to be effective for learning other motor tasks in older adults with cognitive impairment (van Halteren-van Tilborg et al., 2007; Werner et al., 2017).

Finally, higher gestural performances were closely related to higher CRRs at pre- and post-test, supporting our hypothesis and highlighting the high dependence of the robot-integrated GRS on the quality of a user's input for successful gesture recognition. The high extent by which improvements in the gestural performance parallel improvements in the CRR further emphasizes this dependence and suggest that improving the gestural performance of the users is directly translated into improvements in the CRR, leading to an overall improved HRI.

5. Limitations

This study has some limitations. First, the sample size was rather small, limiting the statistical power and generalizability of the results and the ability to perform multiple regression analyses. Second, training effects might not be generalizable to gesture-based HRI with another assistive robot. However, the predefined set of GCs for the assistive bathing robot included various GCs, suggesting that our training approach might be beneficial to improve the gestural performance of older users also for gesture-based HRI with other assistive robots. Third, as a quasi-experimental pretest-posttest study with no control was performed, training effects cannot unequivocally be attributed to the user training. Improvements in the CRR were, however, highly associated with those in the gestural performance, suggesting at least a causal relationship between improving the quality of a user's GCs and improving the CRR and overall HRI, respectively, which was the starting point of this study. Fourth, the study did not include a follow-up, and therefore the sustainability of training effects remains unclear. Future studies should investigate whether potential users are able to remember and correctly perform the GCs also after long periods of time or whether recurrent training sessions are necessary to ensure an adequate gestural performance for long-term successful HRI.

6. Conclusions

The present study reveals that providing a user training specifically tailored to the needs of potential robot users to improve their GCs is highly beneficial for gesture-based HRI with an assistive bathing robot. Our results demonstrated that improved gestural performance is directly translated into better technical performance of the robot-integrated GRS, leading to an overall improved gesture-based HRI. Training benefits can also be achieved in persons with mild-to-moderate cognitive impairment. Older users with low initial gestural performance and more negative feelings toward technology may even benefit the most from a tailored user training. Current findings highlight that for improving gesture-based HRI between assistive robots and older users, future developments and studies in this field should focus not only on refining technical aspects of the robot but also on improving the quality of a user's input by training. Training procedures may be particularly effective when considering the individual resources and limitations of potential users. The presented user training may represent a model for training older adults in gesture-based HRI with an assistive robot.

Funding

This study was supported by the European Commission's Horizon 2020 research and innovation programme (H2020-EU.3.1.4. - Active ageing and self-management of health) under the project "I-SUPPORT - ICT-Supported Bath Robots" (grant agreement no. 643666). The funding source had no role in the design and conduct of the study; collection, analysis and interpretation of data; and preparation, review or approval of the manuscript.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Christian Werner: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Nikos Kardaris:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing. **Petros Koutras:** Software, Data curation, Writing - review & editing. **Athanasia Zlantintsi:** Software, Data curation, Writing - review & editing. **Petros Maragos:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Funding acquisition. **Jürgen M. Bauer:** Resources, Writing - review & editing. **Klaus Hauer:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

None.

Acknowledgements

We kindly thank Michael Günther-Lange and Martin Bongartz for their assistance in participant screening and recruitment. We also thank the volunteers for their willingness to participate in the study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.archger.2019.103996.

References

Afshar, S., & Salah, A. A. (2016). Facial expression recognition in the wild using improved dense trajectories and fisher vector encoding. *2016 IEEE conference on computer vision and pattern recognition workshops (CVPRW),* 1517–1525. https://doi.org/10.1109/CVPRW.2016.189.

Ahluwalia, S. C., Gill, T. M., Baker, D. I., & Fried, T. R. (2010). Perspectives of older persons on bathing and bathing disability: A qualitative study. *Journal of the American Geriatrics Society, 58*, 450–456 10.1111%2Fj.1532-5415.2010.02722.x.

Alameda-Pineda, X., & Horaud, R. (2015). Vision-guided robot hearing. *The International Journal of Robotics Research, 34*, 437–456 10.1177%2F0278364914548050.

Baraldi, L., Paci, F., Serra, G., Benini, L., & Cucchiara, R. (2014). Gesture recognition in egoentric videos using dense trajectories and hand segmentation. *2014 IEEE conference on computer vision and pattern recognition workshops,* 702–707. https://doi.org/10.1109/CVPRW.2014.107.

Begum, M., Wang, R., Huq, R., & Mihailidis, A. (2013). Performance of daily activities by older adults with dementia: The role of an assistive robot. *IEEE international conference on rehabilitation robotics 2013,* 6650405. https://doi.org/10.1109/ICORR.2013.6650405.

Brefka, S., Dallmeier, D., Muhlbauer, V., von Arnim, C. A. F., Bollig, C., Onder, G., et al. (2019). A proposal for the retrospective identification and categorization of older people with functional impairments in scientific studies-recommendations of the medication and quality of life in frail older persons (MedQol) research group. *Journal of the American Medical Directors Association, 20*, 138–146. https://doi.org/10.1016/j.jamda.2018.11.008.

Broadbent, E., Kuo, I. H., Lee, Y. I., Rabindran, J., Kerse, N., Stafford, R., et al. (2010). Attitudes and reactions to a healthcare robot. *Telemedicine Journal and e-Health, 16*, 608–613. https://doi.org/10.1089/tmj.2009.0171.

Caruana, R., Karampatziakis, N., & Yessenalina, A. (2008). An empirical evaluation of supervised learning in high dimensions. *Proceedings of the 25th international conference on machine learning* (pp. 96–103). . https://doi.org/10.1145/1390156.1390169.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 1–27. https://doi.org/10.1145/1961189.1961199.

Chen, K., & Chan, A. H. (2014). Gerontechnology acceptance by elderly Hong Kong Chinese: A senior technology acceptance model (STAM). *Ergonomics, 57*, 635–652. https://doi.org/10.1080/00140139.2014.895855.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: Routledge.

Craik, F. I. M., & Salthouse, T. A. (2008). *The handbook of aging and cognition* (3rd ed.). New York: Psychology Press.

Dalal, N., Triggs, B., & Schmid, C. (2006). *Human detection using oriented histograms of flow and appearance.* Berlin Heidelberg, Berlin, Heidelberg: Springer428–441. https://doi.org/10.1007/11744047_33.

Dawson, D., Hendershot, G., & Fulton, J. (1984). *Aging in the eighties. Functional limitations of individuals age 65 years and over. Advance data from vital and health statistics. Public health services; 1987. DHHS pub. No. 133*.

Delbaere, K., Close, J. C., Mikolaizak, A. S., Sachdev, P. S., Brodaty, H., & Lord, S. R. (2010). The falls efficacy scale international (FES-I). A comprehensive longitudinal validation study. *Age and Ageing, 39*, 210–216. https://doi.org/10.1093/ageing/afp225.

Dews, P. B. (1977). Rate-dependency hypothesis. *Science, 198*, 1182–1183. https://doi.org/10.1126/science.563103.

Dias, N., Kempen, G. I. J. M., Todd, C. J., Beyer, N., Freiberger, E., Piot-Ziegler, C., et al. (2006). Die Deutsche Version der Falls Efficacy Scale-International Version (FES-I). *Zeitschrift für Gerontologie und Geriatrie, 39*, 297–300. https://doi.org/10.1007/s00391-006-0400-8.

Dyck, J. L., & Smither, J. A.-A. (1994). Age differences in computer anxiety: The role of computer experience, gender and education. *Journal of Educational Computing Research, 10*, 239–248 10.2190%2FE79U-VCRC-EL4E-HRYV.

Efthimiou, E., Fotinea, S., Goulas, T., Koutsombogera, M., Karioris, P., Vacalopoulou, A., et al. (2016). The MOBOT rollator human-robot interaction model and user evaluation process. *2016 IEEE symposium series on computational intelligence (SSCI),* 1–8. https://doi.org/10.1109/SSCI.2016.7850061.

El-laithy, R. A., Huang, J., & Yeh, M. (2012). Study on the use of microsoft kinect for robotics applications. *Proceedings of the 2012 IEEE/ION position, location and navigation symposium,* 1280–1288. https://doi.org/10.1109/PLANS.2012.6236985.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146.

Engelhardt, K., & Edwards, R. (1992). Human-robot integration for service robots. In M. Rahimi, & W. Karwowski (Eds.). *Human-robot interaction.* London, UK: Taylor & Francis.

Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., et al. (2016). Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems, 75*, 60–78. https://doi.org/10.1016/j.robot.2014.09.029.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Fong, J. H., Mitchell, O. S., & Koh, B. S. K. (2015). Disaggregating activities of daily living limitations for predicting nursing home admission. *Health Services Research, 50*, 560–578. https://doi.org/10.1111/1475-6773.12235.

Gauggel, S., & Birkner, B. (1999). Validität und Reliabilität einer deutschen Version der geriatrischen Depressionsskala (GDS). *Zeitschrift für Klinische Psychologie und Psychotherapie, 28*, 18–27. https://doi.org/10.1026//0084-5345.28.1.18.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257–283. https://doi.org/10.1146/annurev-psych-113011-143802.

Goodrich, M. A., & Schultz, A. C. (2008). Human–robot interaction: A survey. *Foundations and Trends® in Human–Computer Interaction, 1*, 203–275. https://doi.org/10.1561/1100000005.

Guler, A., Kardaris, N., Chandra, S., Pitsikalis, V., Werner, C., Hauer, K., et al. (2016). *Human joint angle estimation and gesture recognition for assistive robotic vision.* Cham: Springer International Publishing415–431. https://doi.org/10.1007/978-3-319-48881-3_29.

Guralnik, J. M., Ferrucci, L., Simonsick, E. M., Salive, M. E., & Wallace, R. B. (1995). Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *The New England Journal of Medicine, 332*, 556–561. https://doi.org/10.1056/NEJM199503023320902.

Guralnik, J. M., Simonsick, E. M., Ferrucci, L., Glynn, R. J., Berkman, L. F., Blazer, D. G., et al. (1994). A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology, 49*, M85–94. https://doi.org/10.1093/geronj/49.2.M85.

Gure, T. R., Langa, K. M., Fisher, G. G., Piette, J. D., & Plassman, B. L. (2013). Functional limitations in older adults who have cognitive impairment without dementia. *Journal of Geriatric Psychiatry and Neurology, 26*, 78–85. https://doi.org/10.1177/0891988713481264.

Hakkinen, A., Heinonen, M., Kautiainen, H., Huusko, T., Sulkava, R., & Karppi, P. (2007). Effect of cognitive impairment on basic activities of daily living in hip fracture patients: A 1-year follow-up. *Aging Clinical and Experimental Research, 19*, 139–144. https://doi.org/10.1007/BF03324680.

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clinics in Geriatric Medicine, 29*, 737–752. https://doi.org/10.1016/j.cger.2013.07.002.

Haskell, W. L. (1994). J.B. Wolffe memorial lecture. Health consequences of physical activity: Understanding and challenges regarding dose-response. *Medicine and Science in Sports and Exercise, 26*, 649–660. https://doi.org/10.1249/00005768-199406000-00001.

Hauer, K. (2018). Bewertung von AAL-Ambient-Assisted-Living-Systemen bei Personen mit kognitiver Schädigung: Match vs. Mismatch. In O. Bendel (Ed.). *Pflegeroboter* (pp. 89–111). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-22698-5_5.

Hauer, K., Yardley, L., Beyer, N., Kempen, G., Dias, N., Campbell, M., et al. (2010). Validation of the falls efficacy scale and falls efficacy scale international in geriatric patients with and without cognitive impairment: Results of self-report and interview-based questionnaires. *Gerontology, 56*, 190–199. https://doi.org/10.1159/000236027.

Hernandez-Belmonte, U. H., & Ayala-Ramirez, V. (2016). Real-time hand posture recognition for human-robot interaction tasks. *Sensors (Basel, Switzerland), 16*, 36. https://doi.org/10.3390/s16010036.

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise, 41*, 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278.

Huang, L., Zhang, X., & Li, W. (2016). Dense trajectories and dhog for classification of viewpoints from echocardiogram videos. *Computational and Mathematical Methods in Medicine, 2016*, 7. https://doi.org/10.1155/2016/9610192.

Hunter, S. K., Pereira, H. M., & Keenan, K. G. (2016). The aging neuromuscular system and motor performance. *Journal of Applied Physiology, 121*, 982–995. https://doi.org/10.1152/japplphysiol.00475.2016.

International Organization for Standardization (2012). *Robots and robotic devices - vocabulary (ISO 8373:2012(en)*. https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en.

Jagger, C., Arthur, A. J., Spiers, N. A., & Clarke, M. (2001). Patterns of onset of disability in activities of daily living with age. *Journal of the American Geriatrics Society, 49*, 404–409. https://doi.org/10.1046/j.1532-5415.2001.49083.x.

Jones, A., Dawyer, L., Bercovitz, A., & Strahan, G. (2009). The national nursing home survey: 2004 overview. National center for health statistics. *Vital and Health Statistics Series, 13*.

Kardaris, N., Rodomagoulakis, I., Pitsikalis, V., Arvanitakis, A., & Maragos, P. (2016). A platform for building new human-computer interface systems that support online automatic recognition of audio-gestural commands. *Proceedings of the 24th ACM international conference on Multimedia* (pp. 1169–1173). . https://doi.org/10.1145/2964284.2973794.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association, 185*, 914–919. https://doi.org/10.1001/jama.1963.03060120024016.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics, 68*, 355–367. https://doi.org/10.1016/j.ergon.2017.02.004.

Louie, W. Y., McColl, D., & Nejat, G. (2014). Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assistive Technology, 26*, 140–150. https://doi.org/10.1080/10400435.2013.869703.

Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel index. *Maryland State Medical Journal, 14*, 61–65.

Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems, 63*, 22–35. https://doi.org/10.1016/j.robot.2014.09.031.

McNeill, D. (1992). *Hand in mind: Waht gestures reveal about thought.* Chicago, IL: University of Chicago Press.
Naeemabadi, M., Dinesen, B., Andersen, O. K., & Hansen, J. (2018). Investigating the impact of a motion capture system on Microsoft Kinect v2 recordings: A caution for using the technologies together. *PLoS One, 13*, e0204052. https://doi.org/10.1371/journal.pone.0204052.
Nelson, E. A., & Dannefer, D. (1992). Aged heterogeneity: Fact or fiction? The fate of diversity in gerontological research. *The Gerontologist, 32*, 17–23. https://doi.org/10.1093/geront/32.1.17.
Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd international conference on machine learning* (pp. 625–632). . https://doi.org/10.1145/1102351.1102430.
Platt, J. (2000). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schurrmans (Eds.). *Advances in large margin classifiers*. MIT Press.
Pritchard, J. M., Kennedy, C. C., Karampatos, S., Ioannidis, G., Misiaszek, B., Marr, S., et al. (2017). Measuring frailty in clinical practice: A comparison of physical frailty assessment methods in a geriatric out-patient clinic. *BMC Geriatrics, 17*, 264. https://doi.org/10.1186/s12877-017-0623-0.
Rodomagoulakis, I., Kardaris, N., Pitsikalis, V., Mavroudi, E., Katsamanis, A., Tsiami, A., et al. (2016). Multimodal human action recognition in assistive human-robot interaction. *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP),* 2702–2706. https://doi.org/10.1109/ICASSP.2016.7472168.
Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*, 37–59. https://doi.org/10.1300/J079v21n04_02.
Schmidt, L. I., & Wahl, H. W. (2019). Predictors of performance in everyday technology tasks in older adults with and without mild cognitive impairment. *The Gerontologist, 59*, 90–100. https://doi.org/10.1093/geront/gny062.
Schuldt, C., Laptev, I., & Caputo, B. (2004). Recognizing human actions: A local SVM approach. *Proceedings of the 17th international conference on pattern recognition, 2004: Vol. 33*, (pp. 32–36). . https://doi.org/10.1109/ICPR.2004.1334462.
Scopelliti, M., Giuliani, M. V., D'Amico, A. M., & Fornara, F. (2004). If i had a robot at home… peoples' representation of domestic robots. In S. Keates, J. Clarkson, P. Langdon, & P. Robinson (Eds.). *Designing a more inclusive world* (pp. 257–266). London: Springer. https://doi.org/10.1007/978-0-85729-372-5_26.
Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric depression scale (GDS): Recent evidence and development of a shorter version. In T. L. Brink (Ed.). *Clinical gerontology: A guide to assessment and intervention* (pp. 165–173). New York, NY: The Haworth Press.
Smarr, C.-A., Prakash, A., Beer, J. M., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2012). *Older adults' preferences for and acceptance of robot assistance for everyday living tasks. Proceedings of the human factors and ergonomics society … annual meeting. human factors and ergonomics society. annual meeting, 56*, 153–157 10.1177%2F1071181312561009.
Smarr, C. A., Mitzner, T. L., Beer, J. M., Prakash, A., Chen, T. L., Kemp, C. C., et al. (2014). Domestic robots for older adults: Attitudes, preferences, and potential. *International Journal of Social Robotics, 6*, 229–247. https://doi.org/10.1007/s12369-013-0220-0.
Snider, S. E., Quisenberry, A. J., & Bickel, W. K. (2016). Order in the absence of an effect: Identifying rate-dependent relationships. In: *Behavioural Processes, 127*, 18–24. https://doi.org/10.1016/j.beproc.2016.03.012.
Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *International Journal of Social Robotics, 6*, 17–32. https://doi.org/10.1007/s12369-013-0186-y.
Tacken, M., Marcellini, F., Mollenkopf, H., Ruoppila, I., & Széman, Z. (2005). Use and acceptance of new technology by older people. Findings of the international mobilate survey: 'Enhancing mobility in later life'. *Gerontechnology, 3*, 126–137. https://doi.org/10.4017/gt.2005.03.03.002.00.
van Halteren-van Tilborg, I. A., Scherder, E. J., & Hulstijn, W. (2007). Motor-skill learning in Alzheimer's disease: A review with an eye to the clinical practice. *Neuropsychology Review, 17*, 203–212. https://doi.org/10.1007/s11065-007-9030-1.
Vanbellingen, T., Kersten, B., Van Hemelrijk, B., Van de Winckel, A., Bertschi, M., Muri, R., et al. (2010). Comprehensive assessment of gesture production: A new test of upper limb apraxia (tulia). *European Journal of Neurology, 17*, 59–66. https://doi.org/10.1111/j.1468-1331.2009.02741.x.
Veronese, N., Bolzetta, F., Toffanello, E. D., Zambon, S., De Rui, M., Perissinotto, E., et al. (2014). Association between short physical performance battery and falls in older people: The Progetto Veneto Anziani study. *Rejuvenation Research, 17*, 276–284. https://doi.org/10.1089/rej.2013.1491.
Voelcker-Rehage, C., & Willimczik, K. (2006). Motor plasticity in a juggling task in older adults: A developmental study. *Age and Ageing, 35*, 422–427. https://doi.org/10.1093/ageing/afl025.
Wang, H., Kläser, A., Schmid, C., & Liu, C. (2011). Action recognition by dense trajectories. *CVPR,* 3169–3176. https://doi.org/10.1109/CVPR.2011.5995407.
Werle, J., & Hauer, K. (2016). Design of a bath robot system — User definition and user requirements based on international classification of functioning, disability and health (ICF). *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN),* 459–466. https://doi.org/10.1109/ROMAN.2016.7745159.
Werner, C., Ullrich, P., Geravand, M., Peer, A., & Hauer, K. (2016). Evaluation studies of robotic rollators by the user perspective: A systematic review. *Gerontology, 62*, 644–653. https://doi.org/10.1159/000444878.
Werner, C., Wiloth, S., Lemke, N. C., Kronbach, F., Jansen, C. P., Oster, P., et al. (2017). People with dementia can learn compensatory movement maneuvers for the sit-to-stand task: A randomized controlled trial. *Journal of Alzheimer's Disease, 60*, 107–120. https://doi.org/10.3233/JAD-170258.
Wiener, J. M., Hanley, R. J., Clark, R., & Van Nostrand, J. F. (1990). Measuring the activities of daily living: Comparisons across national surveys. *Journal of Gerontology, 45*, S229–237. https://doi.org/10.1093/geronj/45.6.S229.
Wu, Y. H., Wrobel, J., Cornuet, M., Kerherve, H., Damnee, S., & Rigaud, A. S. (2014). Acceptance of an assistive robot in older adults: A mixed-method study of human-robot interaction over a 1-month period in the living lab setting. *Clinical Interventions in Aging, 9*, 801–811. https://doi.org/10.2147/CIA.S56435.
Yamada, K., Yoshida, T., Sumi, K., Habe, H., & Mitsugami, I. (2017). Spatial and temporal segmented dense trajectories for gesture recognition. *SPIE*. https://doi.org/10.1117/12.2266859.
Yang, Y., & Lee, L. C. (2010). Dynamics and heterogeneity in the process of human frailty and aging: Evidence from the U.S. older adult population. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences, 65B*, 246–255. https://doi.org/10.1093/geronb/gbp102.
Zafrani, O., & Nimrod, G. (2018). Towards a holistic approach to studying human–robot interaction in later life. *The Gerontologist, 59*, e26–e36. https://doi.org/10.1093/geront/gny077.
Zhang, J., Marszałek, M., Lazebnik, S., & Schmid, C. (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision, 73*, 213–238. https://doi.org/10.1007/s11263-006-9794-4.
Zlatintsi, A., Rodomagoulakis, I., Koutras, P., Dometios, A. C., Pitsikalis, V., Tzafestas, C. S., et al. (2018). Multimodal signal processing and learning aspects of human-robot interaction for an assistive bathing robot. *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP),* 3171–3175.