International Journal of Social Robotics (2020) 12:21–33
https://doi.org/10.1007/s12369-019-00533-y

The Humanoid Robot NAO as Trainer in a Memory Program for Elderly People with Mild Cognitive Impairment

Olimpia Pino[1] · Giuseppe Palestra[2] · Rosalinda Trevino[1] · Berardina De Carolis[3]

Accepted: 15 February 2019 / Published online: 27 February 2019

Abstract
Many studies on social interaction have used the humanoid robot NAO. In the present paper, we described our project designed to address the growing unmet need for alternative approaches to slowing the progression of cognitive decline in Mild Cognitive Impairment patients. NAO is the experimental platform used in an ecological setting: a center for the treatment of cognitive disorders and dementia of the Italian health service. This paper describes the study addressed to evaluate the effectiveness of human–robot interaction to reinforce therapeutic behavior and treatments adherence and presents the latest findings of functional tests and users investigation recently conducted. The robot was programmed to implement some tasks from the usual memory-training program protocol. In different training conditions, subjects participated in sessions with the support of NAO or only from the psychologist while the interaction was recorded for subsequent exploration. Data indicated that memory training with NAO resulted in an increase of visual gaze from patients and reinforce of therapeutic behavior reducing, in some cases, depressive symptoms. Unexpectedly, significant changes in prose memory and verbal fluency measures were detected. These findings suggest that further research on robotics in ecological settings is necessary to determine the extent to which they can effectively support clinical practice.

Keywords Mild Cognitive Impairment · Social robot · Elderly people

1 Introduction

Combining ICT and the social environment of elderly, to support a widely selfdetermined independent life in their own homes, has led to the development of concepts, products and services. Humanoid robots are able to improve mood, emotional expressiveness and social relationships among patients with dementia [16, 20, 35] also executing many assistive functionalities [15, 17, 19] and providing life assistance demonstrating that the information support provided by the robot also has the potential to improve the daily life of persons with a mild level of dementia [35]. Most recent advances in information and communication technologies have enabled the development of telepresence robots to connect a family member and a person with dementia as a means of enhancing communication between these two parties [23]. The humanoids skills are progressively enhanced: they are able to recognize faces, call people by their name, shape their behavior considering the mood of people interacting with them [7]. Some robots can also reproduce emotions [7, 8], making their human mate feel welcomed, and simulating empathy [6, 10]. Kinetics technology can help them reproduce movements [9], while speech recognition software allows them to respond to what people say, even in many different languages.

During human–robot interaction, the mirror-circuit, responsible for social interaction, is verified to be active [12] suggesting that humans can consider robots as real companions with their own intentions. Considering that, why do not we yet meet robots walking around with people, helping at the supermarket, teaching, assisting the elderly or usually doing any activity in uncontrolled environments? Why has it been so hard for humanoids to leave controlled laboratories and find a place where they can be permanently used? [33]. Many studies have employed the robot NAO. If

✉ Olimpia Pino
olimpia.pino@unipr.it

1 Department of Medicine and Surgery, University of Parma, Parma, Italy

2 Institut des Systmes Intelligents et de Robotique, CNRS UMR7222, Sorbonne Université, Paris, France

3 Department of Computer Science, Universiy of Bari, Bari, Italy

appropriately programmed, it is able to decode human emotions, simulate emotions through the color of its eyes or the position of the body, recognize faces and execute physical exercises with a group of seniors [1], and it is equipped to measure health and environmental parameters [39]. Robotics could partially fill in some of the identified gaps in current healthcare and home care/self-care provisions for promising applications in these fields that we expect to play relevant roles in the near future. With emerging research suggesting that mobile robot systems can improve elderly care [16, 17, 20], also through the development of a coding system aimed at measuring engagement-related behavior across activities in people with dementia [2, 29], it seemed timely to examine whether a humanoid robot could enhance memory performances. Therefore, we evaluated if NAO can assist a memory-training program for people with Mild Cognitive Impairment (MCI) in a center for cognitive disorders therapy, specifically in a memory-training program. MCI refers to a transitional stage between normal aging and early dementia characterized by subjective, objective and heterogeneous decline in cognition, documented by scores below the norm on psychometric tests with preservation of independence in functional abilities [30, 31, 34]. The prevalence of MCI is 10–20% in adults aged 65 years and older. While some MCI patients remain stable or even return to normal over time [11], MCI has a high probability of conversion into dementia at a rate of approximately 10–15% per year. It has been estimated that dementia has been detected at the rate of one new person about every 7 s around the world. Hence, MCI could play a critical role in differencing normal lifespan memory changes from those that are part of disease-related changes. The success in delaying the onset of dementia by as little as a year could reduce the global burden of Alzheimers disease by as much as 9,200,000 cases in 2050, a number that makes every treatment a top priority worldwide [38]. With the purpose of maintaining brain functions, several non-pharmacological interventions are developed. These programs require trained therapists to guide the individual through their performance, to design a new configuration, to provide a useful feedback during the task, and to keep track of the users performance history in order to draw a conclusion on his/her evolution over time [32, 38]. However, space and staff shortages are already becoming an issue, due to an unprecedented increase in life expectancy according to which the global prevalence of cognitive impairment is expected to grow exponentially in the coming years [38]. Our concern in individuals with MCI is to maintain their cognitive capability while they still have their functional abilities and high levels of quality of life and independence [11, 32]. Most of research with NAO has been carried out in controlled environments; thus, it seemed remarkable and valuable to verify if NAO could be used in non-experimental settings introducing it in a typical therapeutic setting using typical protocols with their frequency and intensity. Given the propensity of elderly for engagement with a robot, it was expected that they would enthusiastically respond to one in a healthcare setting [1, 4, 15–17]. It was hypothesized that participants who were supported also by a robot during their memory program would experience higher subjective levels of own mnestic efficiency evaluation, even without cognitive gains. The robot was programmed in order to execute some exercise routines used during the memory training sessions substituting the staff psychologist [9]. Usually memory training has a positive effect on the subjective evaluation of mnestic efficiency. Therefore, in the present study a significant improvement is anticipated in the pertinent scores, while no significant changes are expected in the rest of the neuropsychological tests. It has also been predicted that the humanoid robot could, if appropriately programmed, support a practitioner without significant interferences on the participants attitude and performance. Consequently, no significant differences are expected in the response of the subjects trained with NAO compared with the subjects treated only by the clinician, operationalized as frequency and time length of smiles and glances towards NAO or towards the psychologist while performing the tasks.

In order to analyze equivalent video-clip intervals, training sessions were recorded and examined through a software able to detect faces, measuring frequency and length of the subjects gazes and smiles towards NAO and the practitioner during each task. The exercises were sequentially implemented in three groups of participants allowing us to manipulate NAOs programs as necessary from one group to the subsequent.

2 Methods

2.1 Participants

The participants were selected from the population of outpatients attending the Center for Cognitive Disorders and Dementia of AUSL Parma (Italy), among participants being followed longitudinally across the spectrum of cognitive impairment from December 2015 to February 2017. Here they are involved in programs that last 8 weeks, with weekly meetings of 1 h and a half, conducted in a small group format (6–8 people) by an expert neuropsychologist. All participants were firstly evaluated by memory-disorders specialists and screened. The diagnosis of MCI was based on a detailed medical history, relevant physical and neurological examinations, negative laboratory findings, and neuroimaging studies. For each participant, demographic, clinical, and pharmacological data were formally collected in a detailed case history.

Subjects are enrolled according to the following inclusion criteria: a) diagnosis of MCI obtained through Petersen guidelines, and full marks in the two tests measuring daily living activities (ADL and IADL); b) both genders; c) chronological age comprised between 45 and 85 years; and d) without pharmacological treatment. Exclusion criteria were a diagnosis of major neurocognitive disorder (defined using DSM 5 criteria), history of symptomatic stroke (although silent brain infarction was not an exclusion), history of other central nervous system diseases, serious medical or psychiatric illness that would interfere with study participations, such as Parkinson's disease, HIV/AIDS, or other contraindications. Informed consent was obtained from all the patients or from their legal representatives when appropriate.

The research followed the tenets of the Declaration of Helsinki. Before initiating the robot-guided training program, the experimenter explained to the subjects the study so that they could decide whether to participate. Immediately after the procedures to obtain informed consent, three groups were formed initially of 8 people. Because of absence in several sessions, two subjects of the first group and one of the third group were excluded from the statistical analysis. Finally, the participants were six in the first group, eight in the second group and seven in the last group. The final number of participants was of 21 individuals (10 females and 11 males) with a mean age of 73.45 years (SD = 7.71). The mean education level was of 9.90 years (SD = 4.58) with a minimum value of 5 years corresponding to the conclusion of the elementary school and a maximum value of 18 years which corresponds to a bachelor degree. The gender is well balanced in the sample, even if the first two groups were rather unbalanced, the first with 66.7% of females and the second with only 25%. The average age of the participants is fairly balanced, with the exception of an outlier in the first group under 50 years.

2.2 Statistical Analysis Tool

SPSS 21.0 (IBM Corporation, Armonk, NY, USA) was used for statistical analysis. Group differences in age, years of education, reaction time, error rate data, as well as functional and neuropsychological tests scores were assessed using parametric or non-parametric tests, where appropriate. Chi square test was applied for the analysis of differences in gender. Unpaired t test with Welch correction was applied for values sampled from Gaussian distribution.

2.3 Robotic Platform NAO, Software Architecture and Tools

Figure 1 shows a dialog box using Choregraphe and NaoQi for the Story Reading Task. The Academics version of NAO model H25 (SoftBank Robotics® was used. Some of its

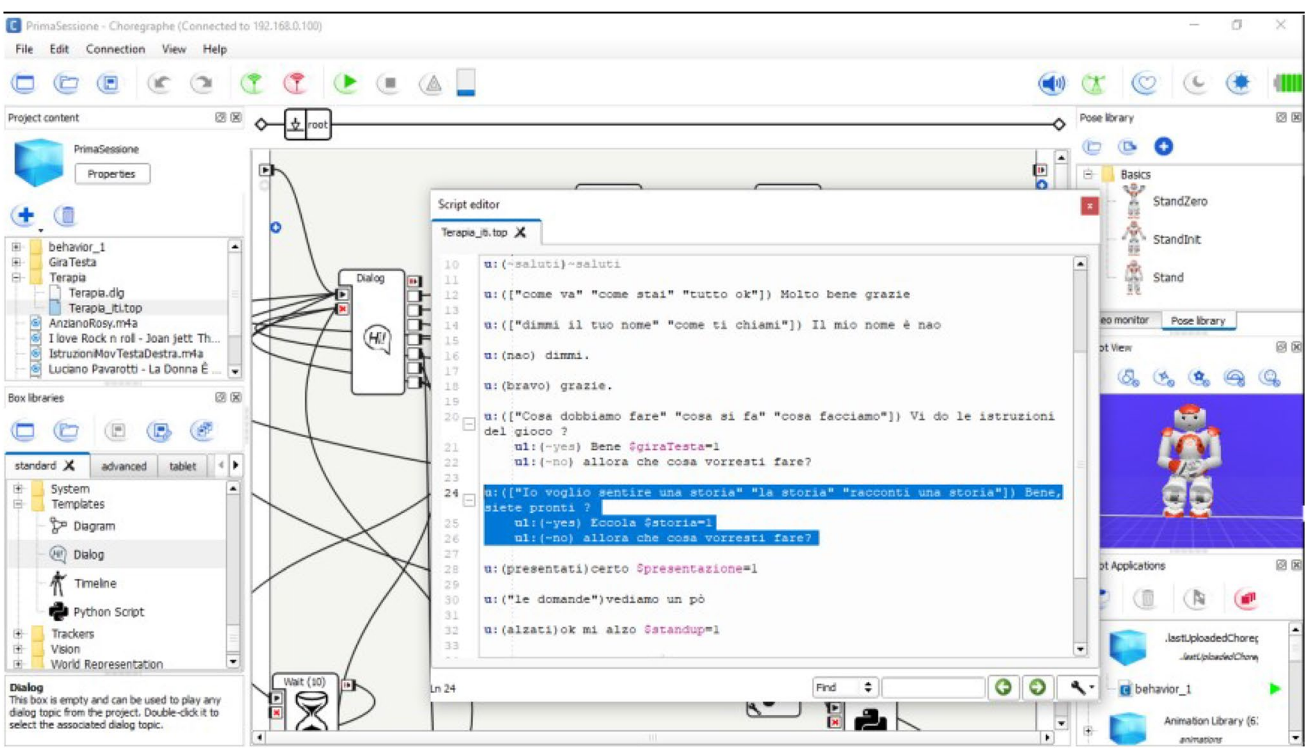

Fig. 1 Dialog box using Choregraphe and NaoQi for the story reading task

features include an on-board fully programmable computer CPU: x86 AMD Geode with 500 MHz, 256 MB SDRAM AND 1 GB flash memory, Wi-Fi and Ethernet connections. It is 58 cm high and weighs 4.3 kg. It uses a 21.6 V rechargeable lithium battery that can keep it running for about 2 h. It has two cameras that can record up to 30 frames/s, 2 hands with self-adapting gripping abilities, but the three fingers of each hand are controlled by a single engine, so they cannot be moved independently. It has force sensitive sensors on hands and feet to perceive contact with objects, light emitting diodes on eyes and body, four microphones to identify the source of sounds and two loud speakers to communicate. It has 25 degrees of freedom in the joints, allowing the movement of head, shoulders, elbows, wrists, firm, waist, legs, knees and ankles independently.

It runs on a native Linux OS platform and it can be programmed through Choregraphe, Python scripts, NaoQi and C++. Our tasks were initially developed using Choregraphe and Python scripts, while NaoQi versions were introduced with the third group of participants. The tasks' versions used with the first two groups included direct voice interactions among NAO and the subjects, leading to a large amount of errors; therefore, the speech recognition software was inhibited for the third group, using instead tactile sensors to simulate an appropriate response of the robot.

2.4 Measures

All individuals underwent to a neuropsychological battery for multiple cognitive domains 1 week before starting the experimental phase (baseline) and at the end of the training (post-treatment). The tests included the following:

(a) Anna Pesenti test to measure episodic memory—verbal Long Term Memory [25]; (b) Digit Span [27] to measure Short Term Memory; (c) Attentional matrices [37] to evaluate visual attention; (d) Memory Assessment Clinics-Questionnaire (MAC-Q, [5]) to measure perceived memory decline with age; (e) Verbal Fluency (PFL, [26]) to assess the ability to access lexicon and lexical organization; (f) Hospital Anxiety and Depression Scale (HADS) [22], to control anxiety and depression levels. In order to determine if the presence of NAO could produce anxiety following the last session with NAO the subjects of all three groups received the STAI-X (State-Trait Anxiety Inventory—[36]). Moreover, other measures were automatically extracted through the analysis of the video recorded sessions. The analysis was made by a customized software that measures smiles and visual attention of the participants. The automatically extracted measures are detailed in the following:

1. Frequency of visual attention (defined as the number of times each patient looks at NAO or the psychologist);
2. Length of visual attention (defined as the time, expressed in seconds, in which each participant turns to NAO or to the psychologist);
3. Frequency of positive expressiveness (defined by the number of times each patient smiles with NAO or with the psychologist);
4. Length of positive expressiveness (defined by time, expressed in seconds, in which a user smiles with NAO or with the clinician).

2.4.1 Impact and Usability of NAO

To evaluate NAO as an assistive tool the following were used: (a) the Psychosocial Impact of Assistive Devices Scales (PIADS—[18]) administered to the three groups to measure NAO's impact on the participants in terms of adaptability to the environment, ability to cope with daily activities and challenges and self-esteem (safety and self-confidence). PIADS evaluates the influence that a device can have in patients using it, measuring along three dimensions. Values range from −3 to +3, with positive values implying a positive change along the specific dimension, and vice versa. A value of 0 indicates that there is no influence; (b) the System Usability Scale (SUS—[3]) measures usability and was only applied to the third group. For this questionnaire, a score of 68 and above indicates that the device is considered as positive and easy to use.

3 Procedure

This section is devoted to the description of the experiment procedure. In order to introduce NAO in the context of a memory training, we analyzed all the exercises that are typically performed during the 8 sessions of the standard program (without the use of a robot). Memory training involved written and verbal practice of memory strategies including visual imagery, association or categorization and spaced retrieval. Volume and duration of training (sessions/week *x* number of weeks) was maintained as the typical format used in the Center. The exercises were extracted from books usually used during the program [13, 14] and aimed to train: (a) focused attention (visual and auditory modalities); (b) divided and alternate attention; (c) categorization and association as learning strategies. We have chosen five tasks to be developed with NAO, considering characteristics that allowed reproducibility using the robot with minimum changes on the exercise. The five selected tasks were:

1. Reading stories;
2. Questions about the story;
3. Associated/not associated words;
4. Associated/not associated word recall;

5. Song-singer match;

Equivalent exercises were also performed by the psychologist in different sessions to obtain comparative data. In both cases, sessions were held in a room where the patients sat around a table. To acquire information on errors made by the patients during the interaction with NAO, two cameras were placed on the longer sides of the table as shown in Fig. 2, with the first two groups.

3.1 First Group

The first of the three groups was selected to represent the pilot study with the aim to develop the tasks that will be performed by NAO. The first version of the exercises was developed using the most common programming method of memory training applications, suitable for PC or tablet, i.e. the participant had a number of attempts to answer the first question; in case of a correct answer within the set number of attempts, we move on to the next question. In case of wrong answer within the maximum number of attempts, the response was marked as incorrect answer and the program moves on to the following question. This way, many types of error are gathered, because when exercises are performed with a PC or a tablet, the answer is given pressing a button, using a mouse or through touch screen eliminating error possibilities. With NAO, on the other hand, since the vocal response is possible, many types of errors can arise, namely:

1. NAO incorrectly understands a correct answer (False Negatives). Falsenegative errors occur when NAO incorrectly understands a correct answer, for example if NAO asks "What is the answer to 5 + 8?" and the patient answers "13" but NAO does not understand the response and records it as incorrect.
2. NAO gives a wrong answer (False Positives). False positive errors arise when NAO catalogs a wrong answer as correct, for example if the question was "What is the answer to 5 + 8?" and the subject replies "10" but NAO records it as correct.
3. The patient's answer occurs before NAO is ready to process the words, that is before the usual "Beep" indicating that the robot is ready to receive an answer.
4. NAO interprets the voice as noise and does not react to the sound (especially if there are several voices simultaneously speaking).

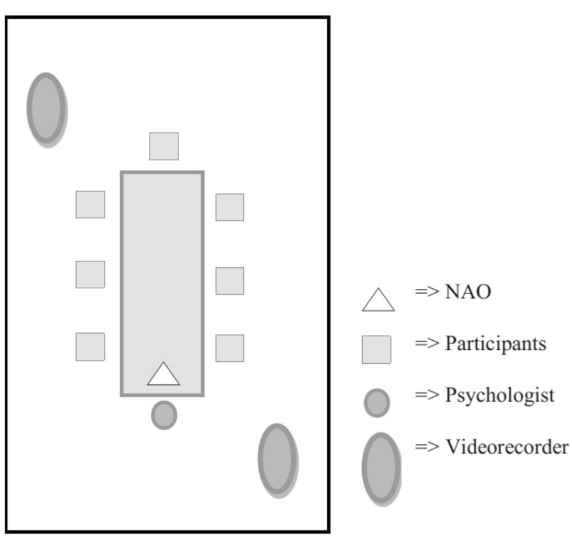

Fig. 2 Layout of the room for the first two groups

3.2 Second Group

In order to capture as much as possible the interaction of participants with NAO robot for the second group, we implemented the same data recording system, i.e. two cameras placed at the corners of the table at a height of about 150 cm from the floor. In this case, the application was modified to avoid automatically handling the question to be asked. It was precisely launched by the operator through the PC that controls NAO. This gains the advantage to ask the question when the subject was ready to answer and being able to manually re-launch the question when an error occurred, without setting a minimum or maximum number of attempts. In order to decide which type of programming could decrease or eliminate errors in the interaction, for the first two groups the amount of errors per type was measured. These exercises were chosen to calculate the number of mistakes (questions on story) because they are more subject to error, since they are open questions and their response can be expressed in multiple ways.

3.3 Third Group

The third group of participants performed the same exercises modified and a new task:

1. Reading a story was adapted replacing NAO's voice with a human voice to facilitate understanding and interaction;
2. Tasks were implemented with Qi-Chat;
3. Patient recognition and naming was added.

To obtain frontal and lateral views for each participant and analyze his/her reactions while interacting with NAO, an additional camera was introduced, as shown in Fig. 3.

In order to decide which programming mode was suitable to reduce or eliminate errors in the interaction, similarly, for the third group the amount of errors per type when answering the questions on the stories spoken by NAO was

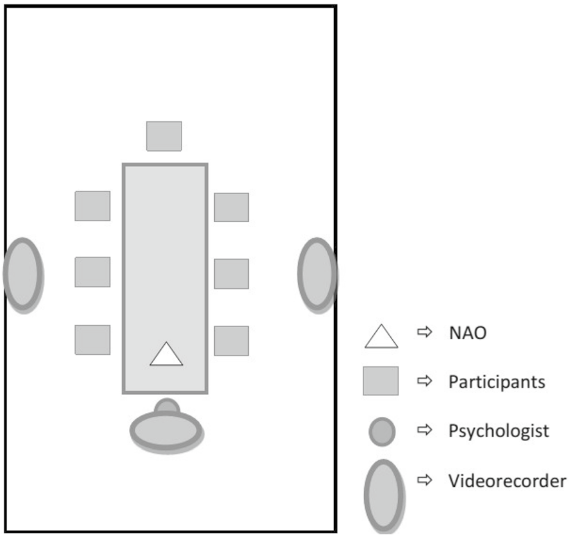

Fig. 3 Layout of the room for the third group

measured. Subjects in this group were administered three other questionnaires during the last meeting with NAO, STAI-X to evaluate state anxiety, PIADS to measure NAO's usefulness as an assistive tool and SUS, to assess NAO's usability.

3.4 Video Analysis Software

The video recordings were cut into fragments of equivalent length and analyzed with a specific and customized software to compare the interaction between the participants and NAO or the psychologist. The video analysis software identifies one or more faces in the video calculating the frequency with which a smile (happy face) appeared and turns towards NAO or towards the psychologist. The video analysis software is based on a previous study that aimed to recognize six basic emotions through facial expression [28]. The original software has been customized in order to detect both happy faces and head pose direction in a specific ecological setting. The ecological environment provides only three cameras for eight participants and specific light conditions.

After identifying a face, the software defines it and some of her/his features (eye contour, lips, eyebrows, nose) locating 77 key points, through a method called Active Shape Model which "fits" in an iterative way a series of points of a face model.

Once the 77 points are identified, the software tracks linear, polygonal, elliptical and angular characteristics (Fig. 4), i.e. the distance between two points to find the following: three lines describing the left eyebrow; two defining the left eye; one for the cheeks; one for nose; eight for the mouth. The software then determines polygonal features, calculating the area delimited by irregular polygons created using three or more key reference points, specifically: one for the left eye; one forming a triangle between the corners of the left eye and the left corner of the mouth; one for the mouth. Thus, the software traces the elliptic characteristics, calculated by the ratio between the major axis and the minor axis of the ellipse, in particular seven ellipses are chosen between the reference points: one for the left eyebrow; three for the eye, left upper and lower eyelid; three for the mouth, lower and upper lips. Once the software has delimited these 32 geometric figures, it will verify if characteristics drawn correspond to a smile or to another expression. To do this, the software uses a classification module that, through a Random Forest classifier, analyzes the geometric characteristic vectors to determine if the participants smiled during the interactions with NAO or with the psychologist.

Additionally, the software allows establishing gaze direction, returning three coordinates in the space and using the tip of the nose to estimate the head pose direction (see Fig. 5). Therefore, knowing NAO's or the psychologist's position, through the intersection of two images taken from

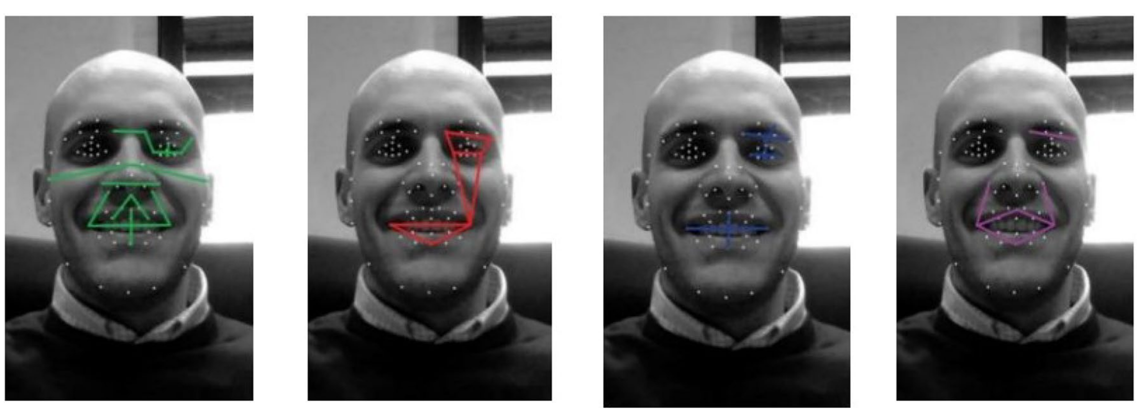

Fig. 4 Linear, polygonal, elliptical and angular characteristics explored by the software in the analysis of human–robot interaction

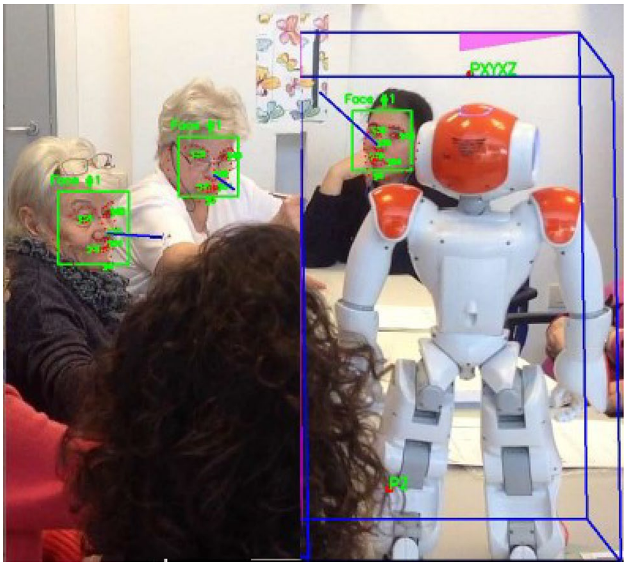

Fig. 5 Visual attention detection based on head pose estimation. The customized software besides detecting smiles, is able to detect the head pose direction towards the robot or the psychologist

the same scene (frontal and lateral), it can be assumed whether the patient is looking at NAO or the psychologist.

The software has been validated on the Extended Cohn-Kanade (CK+) data set, a well known facial expression image database of 123 individuals of different gender, ethnicity and age. It achieves a recognition rate of 97.67% for happiness facial expression. The head pose estimation has been validated on the Radboud Faces Database, a set of pictures of 67 models of different ethnicity, age, and gender displaying emotional expressions using three head orientations (frontal 0, left 45, and right 45). The software reaches an average recognition rate of 98.40% for the three head orientations. Although well known and validated software as Noldus FaceReader [24] and Affectiva [21] are available, the requirements and the constraints of our ecological experimental setting made it necessary to use a customized software solution.

4 Results

4.1 Neuropsychological Evaluations

In Tables 1 and 2 the results of each group for all the neuropsychological measures at T0 and T1 are showed (scores corrected by gender and age). Neuropsychological test scores and cognitive domain scores were approximately normally distributed and summarized as means (M) and standard deviations (SD) in Fig. 6 for Ability, Adaptability and Self-esteem dimensions for the PIADS questionnaire. In Table 3 the frequency of the different types of errors (per group) obtained for NAO or the participants during the interactions on exercises concerned questions on the story is reported. State anxiety (STAI—X) was also measured after the last interaction with NAO, and the mean general score (M = 35.24, SD = 10.43) resulted not far from the one of the single groups (M = 39.33, M = 30.75 and M = 36.86, respectively for the three groups). STAI-X scores from 40 to 50 indicate a state of mild anxiety, and only scores exceeding 60 points are usually considered relevant, therefore, values found are below the minimum level of anxiety immediately after attending a session with NAO, which suggests that the robot does not cause anxiety when used.

The pre- and post-treatment neuropsychological data were analyzed by 3 (groups) x 2 (time periods) ANOVA mixed models. The statistical analysis revealed a significant difference for data of prose memory task, with respect to the time in which measurements were executed [F (1,

Table 1 Anxiety (AN), depression (DE), digit span (SP) and prose (P) mean scores at T0 (baseline) and T1 (Post-treatment measures) (standard deviations in parenthesis)

Group	AN T0	AN T1	DE T0	DE T1	SP T0	SP T1	P T0	P T1
1	8.7 (4.9)	9.2 (5.5)	5.3 (2.2)	6.8 (3.1)	5.2 (0.6)	5.5 (0.6)	6.3 (2.6)	9.5 (3.4)
2	6.0 (2.0)	4.9 (1.3)	4.9 (1.4)	5.4 (1.7)	5.2 (0.8)	5.6 (0.9)	10.1 (5.3)	12.3 (6.4)
3	6.7 (2.7)	6.8 (2.9)	7.4 (2.0)	7.9 (4.3)	5.0 (1.2)	4.8 (1.2)	10.7 (6.8)	13.7 (7.6)

Table 2 MAC-Q, fluency (F), attention (A) mean scores at T0 and T1, and STAI-X at T1 (standard deviations in parenthesis)

Group	MAC-Q T0	MAC-Q T1	F T0	F T1	A T0	A T1	STAI-X
1	26.7 (2.2)	23.3 (4.5)	37.7 (9.4)	36.0 (9.7)	51. 7(4.3)	55.9 (7.6)	39.3 (5.6)
2	25.3 (2.8)	25.6 (1.3)	29.2 (6.5)	34.5 (9.4)	39.02 (9.4)	42.4 (7.3)	30.8 (14.4)
3	27.6 (2.4)	26.6 (3.0)	30.7 (13.3)	36.6 (13.6)	42.4 (10.4)	41.4 (7.5)	36.9 (6.9)

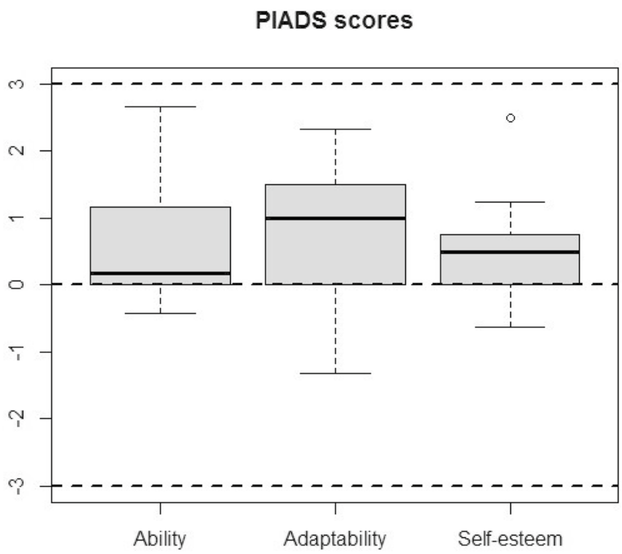

Fig. 6 Mean values for the Psychosocial Impact of Assistive Devices Scales (PIADS)

Table 3 Type of error per group by both NAO and participants in the exercise of listening story and response to questions

Group	No NAO's reaction	Participant error	False negative	False positive	Correct answers
A	6	10	12	5	9
B	13	5	2	1	16
C	0	0	0	0	16

18) = 9.128, $p < 0.007$], for verbal fluency [F (1, 18) = 9.650, $p < 0.006$], indicating that the scores in these tests significantly improved in the post-treatment. A significant interaction between group and time [F (2, 18) = 4.243, $p < 0.030$] emerged for the verbal fluency scores indicating an increase following experimental treatment.

Concerning the evaluation of attention, a significant difference was noted for the interaction between time and group [F (2, 18) = 6.08, $p < 0.009$] suggesting that attention scores vary between the groups according to the measuring period. Tests measuring Anxiety, Depression and Short-Term Memory (Digit Span), as expected, did not show any significant improvement. Respect to PIADS (see Fig. 6), along the Ability dimension, although all groups have evaluated the tool in a positive way, score for the first group appears much dispersed (X = 0.8, SD = 1.1) and with an average below 1 point, while values for the second group are very compact (X = 0.1, SD = 0.3) just above zero. The third group evaluations seems more consistent, considering the device in a positive way (X = 1, SD = 0.6). Also along the Adaptability dimension there is a wide dispersion in judgments of the first group participants (X = 0.94, SD = 1.1), while for the second also negative values are found (X = 0.2, SD = 0.6), suggesting a negative impact in the Adaptability following the experience with NAO. These scores are in contrast with those of the third group (X = 1.1, SD = 1.1), which showed positive scores, all around 1-point. Respect to the Self-esteem dimension, the first group showed closer scores (X = 0.6, SD = 0.97), with an outlier in the positive dimension. The second group participants, with a mean value close to zero (X = 0.2, SD = 0.4), did not perceive a change in Self-esteem and expressed some negative opinions, while the third group provided a positive opinion (X = 0.8, SD = 0.5), with a negative outlier. The System Usability Scale (SUS) provides a usability score of NAO as a device. The mean score found for this scale was of M = 68.57 (SD = 25.32), value at the limit, with a large variability within subjects. It must be considered, however, that the questionnaire was administered only to the third group composed by only seven individuals.

4.2 Session Monitoring: Errors' Type

Regarding the task of listening a story and responding to questions about it, the total number of questions posed by the exercise was of 16. As shown in Table 3, the participants of the first group could not answer all the questions, because once the program cycle was started, it could not be managed, so if NAO committed a false positive error or reached the maximum number of errors set (three), the program led to the following question, leaving the previous one unanswered. With the second group subjects, who received the second version of the program, errors decreased (although NAO committed multiple non-reaction errors), because patients could be guided to respond when NAO was ready to process the answer. This way, all questions could be answered (a total of sixteen). With the third participants group the speech recognition software was inhibited and replaced by the two tactile sensors in NAOs head which, activated by the trainer, threw a positive (if response was correct) or a negative (in case of incorrect response) expressions randomly chosen from two separate lists.

4.3 Participants' Experimental Performances and Human–Robot Interaction Analysis

In Tables 4 and 5 mean values for frequency and time length of both Smiles and Gazes during the different tasks are reported. The story-reading exercise was performed both by NAO and by the psychologist in their standard position: NAO on the table and the psychologist standng by the narrow part of the table, as displayed in Figs. 2 and 3. In both cases, patients listen without reading the story. During the execution of this exercise, NAO did not use its own synthesized voice, but a recorded human voice. As seen in Table 5,

Table 4 Mean values for frequency and time length of smiles during the different tasks (SD in parenthesis)

Tasks	NAO mean (SD)	Trainer mean (SD)	*p* value correct
Story reading			
Frequency	0.7 (0.90)	0.0 (0)	0.38
Time	3.9 (6.20)	0.0 (0)	0.59
Recognition and naming			
Frequency	9.86 (3.07)	4.14 (2.4)	0.02*
Time	170.41 (122.9)	36.24 (24.0)	0.11
Paired words encoding			
Frequency	1.85 (1.34)	0.0 (0)	0.04*
Time	9.70 (11.44)	0.0 (0)	0.26
Paired words recall			
Frequency	2.57 (1.39)	0.43 (0.78)	0.01*
Time	25.83 (25.6)	4.67 (10.29)	0.14
Song-singer matching			
Frequency	5.43 (4.83)	1.43 (1.9)	0.12
Time	30.66 (32.21)	5.01 (6.7)	0.26

Significance levels: * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$

Table 5 Mean values for frequency and time length of visual attention during the different tasks (Standard Deviation in parenthesis)

Tasks	NAO mean (SD)	Trainer mean (SD)	*p* value correct
Story reading			
Frequency	4.0 (2)	3.57 (2.99)	1.00
Time	71.3 (25.42)	74.09 (31.14)	1.00
Recognition and naming			
Frequency	36.0 (7.8)	18.71 (10.1)	0.02*
Time	354.99 (22.3)	157.94 (96.6)	0.007*
Paired words presentation			
Frequency	12.29 (7.48)	11.00 (3.8)	1.00
Time	43.735 (25.95)	60.63 (16.7)	0.015*
Paired words recall			
Frequency	8.57 (4.35)	9.86 (5.27)	1.00
Time	66.24 (33.71)	26.17 (16.96)	0.09
Song-singer matching			
Frequency	23.42 (6.75)	10.14 (5.39)	0.007*
Time	94.51 (43.74)	28.59 (14.54)	0.01*

Significance levels: * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$

no smiles were recorded while the clinician reads the story, while some smiles were recorded when was NAO the reader of the story. Even the frequency with which patients looked at NAO while reading and the length of the gaze is slightly greater, although it did not reach the statistical significativity.

In the patient recognition and naming exercise, the psychologist sat in the typical position around the table, but NAO's position must necessarily vary to allow him to recognize the patients face considering that NAO must have the participant exactly in front to be able to recognize him/her. This peculiarity undoubtedly has an effect on the amount of time that subjects spend directing their gaze towards NAO. Significant comparisons emerged between the frequency of smiles directed to NAO (M = 9.86) and those focused on the psychologist (M = 4.14) [t (12) = 4.382, $p < 0.02$], between the frequencies of the visual interaction towards NAO (M = 35) and that devoted to the psychologist (M = 18.7) [t (12) = 3.99, $p < 0.02$] and, finally, that concerning the length of the visual interaction with NAO (M = 354.99) compared to that on the psychologist (M = 157.9), [t (12) = 5.33, $p < 0.007$]. These data can be considered dependent mostly on an artifact of the procedure.

During the presentation of paired words exercise with the psychologist, patients hold a list of couples of words that she/he must learn by association, so at this stage, they practically do not look at the psychologist, but their own sheet, so the number of times they direct her/his gaze at her and/or smile is really low, or null. NAO, on the contrary, assists them to create the association because the participants hold the sheet in their hand reading the first word and NAO expresses the associated word adding some comments as strategy to highlight each association. This procedural peculiarity can be causing the preeminence of the frequency of smiles towards NAO (M = 1.85) rather than those addressed to the psychologist (M = 0) [t (12) = 3.65, $p < 0.04$], it seems curious that the gaze towards NAO (M = 43.73) lasts significantly less than the one towards the psychologist (M = 60.63) [t (12) = 4.53, $p < 0.015$].

The paired words recall task with the clinician follows practically the same procedure, with a sheet reporting the first word of the paired words and only the stem (the first letter) of the associated word. The patient must recall and write down the associated word presented before. With the humanoid, the input is auditory: NAO pronounces the stimulus of the couple and the participants, one at a time, are requested to replay with the response. Following the answer, NAO delivers an informative feedback. With this manipulation, the frequency for smiles towards NAO (M = 2.57) significantly increases [t (12) = 5.30, $p < 0.007$] compared to those addressed to the psychologist (M = 0.43).

The song-singer matching is a task in which participants have to remember the songs title as response to the name of the singer who made that piece famous. The procedure with the psychologist is performed matching the title with a written response about the correct singer. In the procedure for the experimental condition, instead, NAO sings the song with the original singers voice, waiting for a spoken response (the name of the singer) from the participants and consequently delivering the feedback about its accuracy. In this condition a significant difference was noted in the frequency with which subjects direct their gaze towards NAO (M = 23.42) compared to the looks frequency directed to the trainer (M = 10.14) [t (12) = 5.36, $p < 0.007$] as well as in the gazes length [M = 94.51 vs. M = 28.59, respectively, t (12) = 4.58, $p < 0.01$].

5 Discussion and Conclusions

In the present study, a humanoid robot was used as support in a memory- training program addressed to individuals with MCI. Usually memory programs obtain a positive effect on the subjective self-evaluation of memory efficiency; thus, a significant change was expected on this measure, while no significant fluctuations were expected in the neuropsychological tests. Moreover, it has also been hypothesized that elderly who were sustained by NAO during the training would experience higher levels of subjective memory self-evaluation and lower levels of anxiety than elderly without such presence would. Therefore, no significant differences were expected in their reaction at the tasks while interacting with NAO or the psychologist, measured as frequency and time length of smiles and gazes towards NAO or towards the trainer. On the contrary, we failed to find a significant change in the MAC-Q score, possibly due to the heterogeneity of the sample, embracing individuals affected by all types of MCI and not only the amnesic type. State anxiety level measured following a session with NAO, as expected, exhibited an average value below the mild anxiety threshold. On the other hand, no significant changes were anticipated in the other measures. Unexpectedly, significant alterations were revealed from prose memory and verbal fluency measures. It is impossible to determine whether these differences are due to the presence of NAO, because of the nature of the research design; it would be interesting to extend the procedure to a larger sample of patients distributed in independent groups. The significant change found in attention measures for the second group is probably due to an unknown factor. The assistive tool evaluation, in terms of ability to perform actions, adaptability to the environment and self-esteem, produced an encouraging score. In all the three dimensions, the third group showed a higher score, while the second group recorded a lower mean value. The System Usability Scale qualified NAO as an easy-to-use device. For both questionnaires (PIADS and SUS), patients qualify the instrument usability as the combination of humanoid and relevant programming, since the robot was in some cases managed directly by the operator. Errors accomplished both by NAO and participants during the implementation of the two open question tasks were considered, because that sort of task encompasses a complex interaction. Our findings showed that when the speech recognition software was inhibited, errors disappeared. As expected, no significant differences emerged for patients behavior with the reading-story task; in both contingencies (NAO vs. psychologist) there is an acoustic stimulus; moreover, in the third group NAO's voice had been replaced by a human voice, to avoid using the synthesized voice. Such manipulation ensured that both interactors become more comparable. The significant differences found in the recognizing and naming exercise may be attributed to the required adaptations of the task to make it suitable for NAO. In this case, NAO's facial recognition function was implemented, and to make it work better, NAO had to be placed in front of each patient, while the psychologist performed the whole exercise from one end of the table. The paired words task shows differences in the frequency with which the subjects smile or address their look towards NAO; it is possible to believe that it is due to differences in the execution mode. It should be noted that the exercise version accomplished by the psychologist is performed with paper and pencil; whereas, with NAO the stimulus was not visual, but acoustic. Likewise, the significant difference found in the song-singer matching task maybe due consistently to the same effect. During all the time spent with NAO in the center, much attention has been paid to which tasks to perform with NAO, but the way in which these can be done may differ from one developer to the other. A program can always be analyzed, evaluated or enriched to be re-proposed. A humanoid robot provides engaging situations and, in some circumstances, enthusiastic behaviors were detected in patients as a reaction to some reinforcement phrases after a task, rather than during the task itself, as long as the reinforcement expressions were not repetitive, but casually chosen from a list of general reinforcements. It would be exciting to personalize these sentences with the name of the participant who replied correctly. It would seem like an easy thing to do, but it is necessary to remember that NAO recognizes people by looking at their faces, so that the same tasks should be rewritten. In our layout, participants can sit in a fixed place during the sessions, and take advantage of the patients position with respect to NAO to recognize and reinforce them in a personalized way. This redesign, readjustment and re-presentation of the exercises should be completed cyclically, observing and using feedback from the presentation of a task to make it increasingly variable and close to human behavior, until NAO will be able to move independently without the programmers intervention. NAO's behavior variability is essential with the aim to enhance the interaction with patients, so it would be appropriate to initially arrange diverse versions of the same task, with different movements, attitudes and voice pitch, and then choose randomly between the programs or between program portions, in order to always have a minimum variability in the implementation of the same task. This could deliver the typical freshness of human beings when reading the same sentence twice or supplying the same explanation several times without being exactly the same as the previous one. NAO's programming is usually not considered during the time it does not perform an exercise, for example, when it arrives in the center. Typically, the programmer carries the robot in a suitcase, pulls it out inanimate and then turns on routers, PCs, and robot; this obviously could not be done with a living being. To perceive it as a living body with which it is possible to interact, would not it be better to let it enter the room while walking? Moreover, while NAO is waiting between one task and the other, keeping it motionless in a corner causes the same perception in patients that it would cause if it moves around the room and asks a few questions or makes requests? It would be convenient to observe and improve NAO's behavior and its interactions with the patient in free situations (where the humanoid does not perform any exercise) limiting the movement environment in order to create a space in which it can be self-sufficient once switched on, where it knows what to do and what not to do, that can be quieted with a voice command, but can also give the impression of having the need to move and having its own intentions. This part of the work is a matter of creativity, and should be done according an interdisciplinary perspective, where the work of those who develop and manage the training procedure and study human behavior is complemented by creative engineers who write and imagine variable scripts for the robot that can be launched in specific environments. To our knowledge, no study reported the application of a humanoid robot in the health care setting for individuals suffering of MCI. Our approach is a controlled study which determined that when training is assisted by a robot, elderly with MCI experience more attention and less depressive symptoms during a common memory-training protocol. In addition to the psychological intervention, our investigation suggests that technologically enhanced forms of involvement for the management of non-pharmacological approaches should be considered. If programmed to execute psychological strategies and training, NAO appears promising because it provides a highly engaging arrangement. Moreover, the vast majority of participants indicated they would like to have the robot at home. Results must be understood according to the following limitations. First, the study was carried out in a single center with an experienced psychologist who was interested in and not blinded to the purpose of the study (not possible considering that the robot was talking). Data from multiple centers with different staff members are needed to determine if our results are generalizable. The small number of participants limits this study; however, the focus of the research was on feasibility rather than generalizability. Nevertheless, the use of humanoids in health care setting suggests that robotics can be one of the most important and cost-effective technologies to enter the health care system. For future work, we plan to improve the video software analysis handling facial expression and gaze measures in the context of the events that cause them as described elsewhere [37]. Having determined that this innovative approach has a benefit to patients, it is suitable to next compare it across individuals with varying characteristics such as inpatient versus out-patient status, and so on. Further research is required to explore the role of humanoid robotics but our study offers promising evidence that it can provide a beneficial impact.

Acknowledgements Authors would like to thank the participants who gave their time generously to the study, the clinical teams who supported recruitment to the project, in particular Beatrice Pagliara from the Center of Cognitive Impairments and Dementia of Parma, Italy.

Author's Contribution OP and RT cooperated and contributed to the design and plan of the present study. RT was in charge for data acquisition, analysis and manuscript writing. GP and BDC was in charge of software assessments and analysis. OP and BDC were in charge of manuscript verifying.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bäck I, Makela K, Kallio J (2013) Robot-guided exercise program for the rehabilitation of older nursing home residents. Annal Long Term Care 21(6):38–41
2. Boissy P, Corriveau H, Michaud F, Labonté D, Royer MP (2007) A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects. J Telemed Telecare 13(2):79–84
3. Brooke, J.: Sus: A quick and dirty usability scale. In: T. B., W. A., M.I. I. (eds.) JordanP. W., pp. 189–194 (1996)
4. Cohen-Mansfield J, Dakheel-Ali M, Marx MS (2009) Engagement in persons with dementia: the concept and its measurement. Am J Geriatr Psychiatry 17(4):299–307. https://doi.org/10.1258/135763307780096195
5. Crook TH, Feher EP, Larrabee GJ (1992) Assessment of memory complaint in age associated memory impairment: the mac-q. Int Psychogeriatr 4(2):165–176
6. De Carolis B, Ferilli S, Palestra G (2015) Improving speech-based human robot interaction with emotion recognition. In: International symposium on methodologies for intelligent systems, Springer, pp 273–279
7. De Carolis B, Ferilli S, Palestra G (2017) Simulating empathic behavior in a social assistive robot. Multimed Tools Appl 76(4):5073–5094. https://doi.org/10.1007/s11042-016-3797-0
8. De Carolis BN, Ferilli S, Palestra G, Carofiglio V. Towards an empathic social robot for ambient assisted living. In:

ESSEM@ AAMAS, pp 19–34. URL https://dl.acm.org/citation.cfm?id=3054107
9. Ertuğrul BS, Gurpinar C, Kivrak H, Kose H (2013) Gesture recognition for humanoid assisted interactive sign language tutoring. In: Signal processing and communications applications conference (SIU), pp 1–4
10. Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. Trends Cognitive Sci 8(9):396–403. https://doi.org/10.1016/j.tics.2004.07.002
11. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H et al (2006) Mild cognitive impairment. Lancet 367(9518):1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5
12. Gazzola V, Rizzolatti G, Wicker B, Keysers C (2007) The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. Neuroimage 35(4):1674–1684. https://doi.org/10.1016/j.neuroimage.2007.02.003
13. Gollin D, Ferrari A, Peruzzi A (2007) Una palestra per la mente. stimolazione cognitiva per l'invecchiamento cerebrale e le demenze. Edizioni Erickson, Trento
14. Gollin D, Ferrari A, Peruzzi A (2011) Una palestra per la mente. Erickson Ed, Trento
15. Graf B, Parlitz C, Hägele M (2009) Robotic home assistant care-o-bot® 3 product vision and innovation platform. In: International conference on human–computer interaction, Springer, pp 312–320
16. Gross HM, Schroeter C, Mueller S, Volkhardt M, Einhorn E, Bley A, Langner T, Martin C, Merten M (2011) I'll keep an eye on you: Home robot companion for elderly people with cognitive impairment. In: Systems, man, and cybernetics (SMC), 2011 IEEE international conference on, pp 2481–2488
17. Gross HM, Schroeter C, Mueller S, Volkhardt M, Einhorn E, Bley A, LangnerT, Merten M, Huijnen C, van den Heuvel H, et al. (2012) Further progress towards a home robot companion for people with mild cognitive impairment. In: Systems, man, and cybernetics (SMC), 2012 IEEE international conference on, pp 637–644
18. Jutai J, Day H (2002) Psychosocial impact of assistive devices scale (PIADS). Technol Disabil 14(10):107–111
19. Kidd CD, Breazeal C (2008) Robots at home: understanding long-term human–robot interaction. In: Intelligent robots and systems, IROS 2008. IEEE/RSJ international conference on, pp 3230–3235
20. Martín F, Agüero C, Cañas JM, Abella G, Benítez R, Rivero S, Valenti M, Martínez-Martín P (2013) Robots in therapy for dementia patients. J Phys Agents 7(1):48–55
21. McDuff D, Mahmoud A, Mavadati M, Amr M, Turcot J, Kaliouby Re (2016) Affdexsdk: a cross-platform real-time multi-face expression recognition toolkit. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, pp 3723–3726. ACM . https://doi.org/10.1145/2851581.2890247
22. Montazeri A, Vahdaninia M, Ebrahimi M, Jarvandi S (2003) The hospital anxiety and depression scale (hads): translation and validation study of the iranian version. Health Quality Outcomes 1(1):14
23. Moyle W, Jones C, Cooke M, O'Dwyer S, Sung B, Drummond S (2014) Connecting the person with dementia and family: a feasibility study of a telepresence robot. BMC Geriatr 14(1):7. https://doi.org/10.1186/1471-2318-14-7
24. Noldus (2014) Facereader: tool for automatic analysis of facial expression: Version 6.0
25. Novelli G, Papagno C, Capitani E, Laiacona M (1986) Tre test clinici di ricerca e produzione lessicale. Taratura su sogetti normali. Archivio di Psicologia, Neurologia e Psichiatria. 477–506
26. Novelli P, Capitani L, Vallar C, Cappa S (1986) Test di fluenza verbale. Archivio di Psicologia, Neurologia e Psichiatria **47**(4)
27. Orsini A, Grossi D, Capitani E, Laiacona M, Papagno C, Vallar G (1987) Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. Italian J Neurolo Sci 8(6):537–548
28. Palestra G, Pettinicchio A, Del Coco M, Carcagn`ı P, Leo M, Distante C (2015) Improved performance in facial expression recognition using 32 geometric features. In: International conference on image analysis and processing, Springer, pp 518–528
29. Perugia G, van Berkel R, D´ıaz-Boladeras M, Catal`a-Mallofr´e A, Rauterberg M, Barakova E (2018) Understanding engagement in dementia through behavior. The ethographic and laban-inspired coding system of engagement (elicse) and the evidence-based model of engagement-related behavior (emodeb). Frontiers in psychology **9** https://doi.org/10.3389/fpsyg.2018.00690
30. Petersen RC (2007) Mild cognitive impairment: current research and clinical implications. In: Seminars in neurology, vol 27. Copyright© 2007 by Thieme Medical Publishers, Inc, New York, pp 022–031
31. Pino O (2015) Memory impairments and rehabilitation: evidence-based effects of approaches and training programs. Open Rehabilit J **8**(1)
32. Pino O (2017) Ricucire i ricordi: la memoria, i suoi disturbi, le evidenze di efficacia dei trattamenti riabilitativi. Mondadori universit`a
33. Pino O, Cadena R (2018) Will humanoid robots become an assistive technology for people with special needs. J Alzheimers Dement 1(001):2. https://doi.org/10.21767/IPAD.100001
34. Pino O, Poletti F, Caffarra P (2013) Cognitive demand and reminders effect on time-based prospective memory in amnesic mild cognitive impairment (amci) and in healthy elderly. Open J Med Psychol 2(01):35. https://doi.org/10.4236/ojmp.2013.21007
35. Sabanovic S, Bennett CC, Chang WL, Huber L (2013) Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In: Rehabilitation robotics (ICORR), 2013 IEEE international conference on, pp 1–6
36. Spielberger CD (1983) State-trait anxiety inventory for adults. PsycTESTS Dataset https://doi.org/10.1037/t06496-000
37. Spinnler H (1987) Standardizzazione e taratura italiana di test neuropsicologici. Ital J NeurolSci 6:21–120
38. Ton TG, DeLeire T, May SG, Hou N, Tebeka MG, Chen E, Chodosh J (2017) The financial burden and health care utilization patterns associated with amnestic mild cognitive impairment. Alzheimer's Dementia 13(3):217–224. https://doi.org/10.1038/s41598-018-21983-w
39. Vital JP, Couceiro MS, Rodrigues NM, Figueiredo CM, Ferreira NM (2013) Fostering the nao platform as an elderly care robot. In: 2013 IEEE 2nd international conference on serious games and applications for health (SeGAH), pp 1–5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Olimpia Pino is professor of General Psychology at University of Parma where she is head of the Laboratory of Cognitive Psychology. Editor of *BAOJ Medical and Nursing*, her research activity is characterized by a strong interest for the relationship between psychology and neuroscience. Her current research fields concern the application of Brain Computer Interface for regulation of brainwaves patterns within a few of personality and behavior disorders and the use of the humanoid robot NAO-H25 for memory rehabilitation in individuals with mild cognitive impairment.

Giuseppe Palestra is research associate at the Institute for Intelligent Systems and Robotics (CNRS UMR 7222), Sorbonne Université, Paris, France. He received the PhD degree in Computer Science in 2017 from the University of Bari. His research focuses on Artificial Intelligence, Social Robotics, and Computer Vision. He worked on various smart health projects in Europe. They were concerned with technologies to improve people's lives, including the SARACEN project (Socially Assistive Robots for Autistic Children Education), of which he was the principal investigator. Since then, he has been working on social robotics, robots that are useful to people, and AI.

Rosalinda Trevino has firstly graduated in Chemical Engineering at the ITESM, where she also obtained a Master in Business Management. Wanting to give a turn to her professional life, she obtained a degree in Psychology and completed a master's degree in Psychobiology and Cognitive Neuroscience at the University of Parma.

Berardina De Carolis is Assistant Professor at the Department of Computer Science at the University of Bari, Italy. Her research interests are mainly in Affective Computing and Intelligent Interfaces.