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

International Journal of Social Robotics

The humanoid robot Nao-H25 as trainer in a memory program for elderly people with Mild Cognitive Impairment (MCI)

--Manuscript Draft--

Manuscript Number:	SORO-D-18-00115
Full Title:	The humanoid robot Nao-H25 as trainer in a memory program for elderly people with Mild Cognitive Impairment (MCI)
Article Type:	Original research
Keywords:	Mild Cognitive Impairment; social robot; elderly people
Corresponding Author:	Olimpia Pino, Phd Universita degli Studi di Parma Dipartimento di Medicina e Chirurgia ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita degli Studi di Parma Dipartimento di Medicina e Chirurgia
Corresponding Author's Secondary Institution:	
First Author:	Olimpia Pino, Phd
First Author Secondary Information:	
Order of Authors:	Olimpia Pino, Phd
	Giuseppe Palestra
	Rosalinda Trevino
	Berardina De Carolis
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Many studies on social interaction have used the humanoid robot Nao-H25. In this paper, we introduced Nao in a typical therapeutic environment: a center for the treatment of cognitive disorders. The robot was programmed to execute some exercise used during memory-training program for elderly people with Mild Cognitive Impairment (MCI). In order to analyze training sessions a software able to detect faces, measuring frequency and length of the subjects' gazes, and smiles towards Nao and the practitioner was implemented. No significant differences were expected in patients performances during the exercises while interacting with Nao or the psychologist. Unexpectedly, significant changes were detected in prose memory and verbal fluency measures.

Noname manuscript No.
(will be inserted by the editor)

The humanoid robot Nao-H25 as trainer in a memory program for elderly people with Mild Cognitive Impairment (MCI)

Olimpia Pino · Giuseppe Palestra · Rosalinda Trevino · Berardina De Carolis

Received: date / Accepted: date

Abstract Many studies on social interaction have used the humanoid robot Nao-H25. In this paper, we introduced Nao in a typical therapeutic environment: a center for the treatment of cognitive disorders. The robot was programmed to execute some exercise used during memory-training program for elderly people with Mild Cognitive Impairment (MCI). In order to analyze training sessions a software able to detect faces, measuring frequency and length of the subjects' gazes, and smiles towards Nao and the practitioner was implemented. No significant differences were expected in patients performances during the exercises while interacting with Nao or the psychologist. Unexpectedly, significant changes were detected in prose memory and verbal fluency measures.

Keywords mild cognitive impairment · social robot · elderly people

1 Introduction

Social robots are able to improve mood, emotional expressiveness and social relationships among patients with dementia [14, 21]. They are also used in the autistic spectrum children treatment to improve their social skills, shared attention, turn-taking and understanding other's emotions [2, 16]. In classes and

O. Pino and R. Trevino
Department of Medicine and Surgery
University of Parma (Italy) E-mail: olimpia.pino@unipr.it

Giuseppe Palestra
Institut des Systmes Intelligents et de Robotique
CNRS UMR7222, Sorbonne Universit, Paris (France)

Berardina De Carolis
Department of Computer Science
Universiy of Bari (Italy)

school environments they can be used to foster learning [23,4] and problem-solving skills [20] of mathematics or science [9]. They can also assist users at airports and supermarkets [36] or they can keep company at home [19, 13]. Their skills are progressively enhanced: they are able to recognize faces, call people by their name, read basic emotions and, consequently, shape their behavior considering the mood of people interacting with them [6]. Some of these humanoids can also simulate emotions, making their human mate feel welcomed, simulating empathy through the activation of the mirror neuron circuit [10] of their human companions. Kinetics technology can help them imitate movements [8], while speech recognition software allows them to respond to what people say, even in many different languages. It has been shown that the mirror circuit, responsible for social interaction, is also active during human-robot interaction [12]; this suggests that humans are able to consider them as real companions with their own intentions. If technology and research are at this point, why do not we yet see robots walking around with people, helping in the supermarket, teaching, assisting the elderly or usually doing any activity in uncontrolled environments? Why has it been so difficult for humanoid robots to leave the controlled laboratory environments and find a place where they can be permanently used? [26]. Many studies on social interaction have used the humanoid robot Nao-H25. If appropriately programmed, it is able to understand human emotions, simulate emotions through the color of his eyes or the position of the body, recognize faces, make exercise a group of seniors or engage the attention of autistic children and many other activities. 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 improve elderly care [15], it seemed timely to examine whether a humanoid robot could enhance memory performances. Therefore, we introduced Nao in a center for the treatment of cognitive disorders, specifically in a memory-training program for people with Mild Cognitive Impairment (MCI). Mild cognitive impairment (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 [30,32]. 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 evolving toward dementia at a rate of approximately 10-15% per year; hence, MCI could play a critical role in differenting normal lifespan memory changes from those that are part of disease-related changes. With the purpose of maintaining brain functions, several non-pharmacological interventions are developed. These programs require trained therapists to guide the individual through their execution, to design a new configuration, to provide an useful feedback during the task, and to keep track of the user's performance history in order to draw a conclusion on his/her evolution over time [31]. 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 [35]. Our interest in MCI is to maintain the mental capacity of individuals while they still have their functional abilities and high levels of quality of life and autonomy [27]. Many of the previous research that included Nao have been performed in controlled environments with specifically designed procedures; thus, it seemed stimulating and useful to verify if Nao could be used in non-experimental settings introducing it in a typical therapeutic environment and observing as much as possible the usual procedure. Given the propensity of elderly for engagement with a robot, it was expected that they would enthusiastically respond to one in a healthcare setting [3]. Thus, it was hypothesized that participants who were guided by a robot during their memory program would experience higher subjective levels of their mnestic efficiency evaluation, even without cognitive gains. The robot was programmed to execute some exercise routines used during the memory program substituting the staff psychologist. Usually a memory training has a positive effect on the subjective evaluation of mnestic efficiency. Therefore, in the present study a significant improvement is expected in the relevant scores, while no significant changes are expected in the rest of the neuropsychological tests. It has also been anticipated that a 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 reaction of the subjects trained with Nao compared with the psychologist, operationalized as frequency and time length of smiles and glances towards Nao or towards the psychologist while performing the exercises. 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 exercise. The exercises were sequentially implemented in three groups of patients allowing us to manipulate Nao's programs as necessary from one group to the following.

2 Methods

2.1 Participants

The participants were selected from the population of outpatients attending the Center for Cognitive Disorders and Dementia of Parma (Italy), from 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 hour and a half, conducted in small groups (6-8 people) by an expert neuropsychologist. All participants were firstly evaluated by memory-disorders specialists and screened and 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 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 exercise program, the experimenter explained to the subjects the procedure of 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, but two subjects of the first group and one of the third group were eliminated from the subsequent analysis because they were absent on several occasions. Finally, the participants were six in the first group, eight in the second group and seven in the last group.

2.2 Measures

All individuals underwent to a neuropsychological battery for multiple cognitive domains one 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 [24]; b) Digit Span [28] to measure Short Term Memory; c) Attentional matrices [34] to evaluate visual attention; d) Memory Assessment Clinics-Questionnaire (Mac-Q, [5]) to measure perceived memory decline with age; e) Verbal Fluency (PFL, [25]) 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 (State-Trait Anxiety Inventory - [33]).

2.3 Robotic Platform Nao H25, software architecture and tools

Figure 1 shows a dialog box using Choregraphe and NaoQi for the Story Reading Task. The Academics version of Nao (SoftBank Robotics) was used. Some of its features include an on-board fully programmable computer CPU: x86 AMD Geode with 500 Mhz, 256 MB SDRAM AND 1 GB flash memory, WiFi 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 hours. It

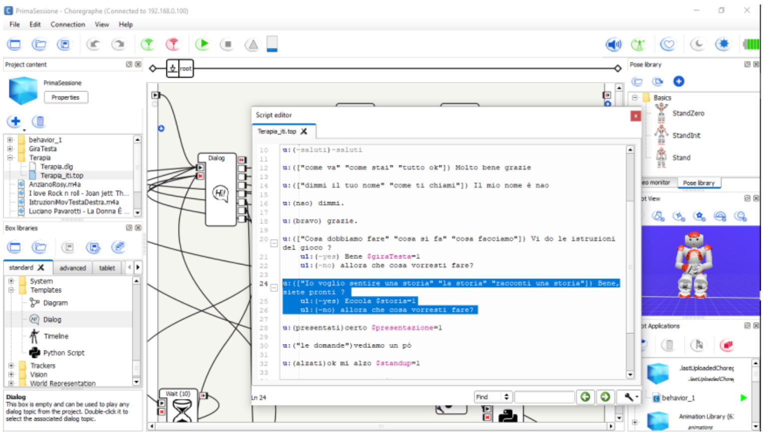

Fig. 1 Dialog box using Choregraphe and NaoQi for the story reading task.

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 Pyton 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 Nao.

2.3.1 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 (see Table 3). 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 - [17]) measures usability and was only applied to the third group. For this questionnaire, a

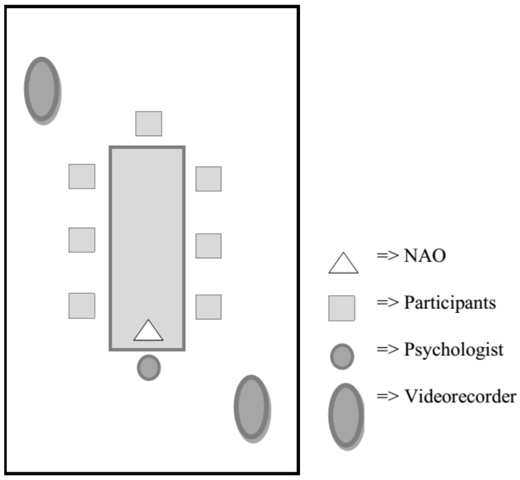

Fig. 2 Layout of the room for the first two groups.

score of 68 and above indicates that the device is considered as positive and easy to use.

2.4 Procedure

Figures 2-3 show the room's layout for the first two groups and the third respectively. Figure 4 includes the facial expression analyzing software flow diagram. Figure 5 shows linear, polygonal, elliptical and angular facial characteristics. In order to introduce Nao in the context of a memory training, we analyzed all the exercises that are tipically performed during the 8 sessions of the program, and we have chosen five of them to be developed with Nao. The exercises were extracted from books usually used during the program [7, 1] and aimed to train: a) focused attention (visual and auditory modalities); b) divided and alternate attention; c) categorization and association as learning strategies. Those developed with Nao were selected considering characteristics that allowed reproducibility using the robot with minimum changes on the excercise (Reading stories, Questions about the story, Associated / not associated words, Associated/not associated word recall, Song-singer match). Equivalent exercises were also performed by the psychologist in different sessions to obtain comparative data.

The 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 Figure 2, with the first two groups. The first of the three groups was selected to represent the pilot study with the aim to develop the exercises performed by Nao. The first version of the exercises was developed using the most common programming method of memory training applications, suitable for PC

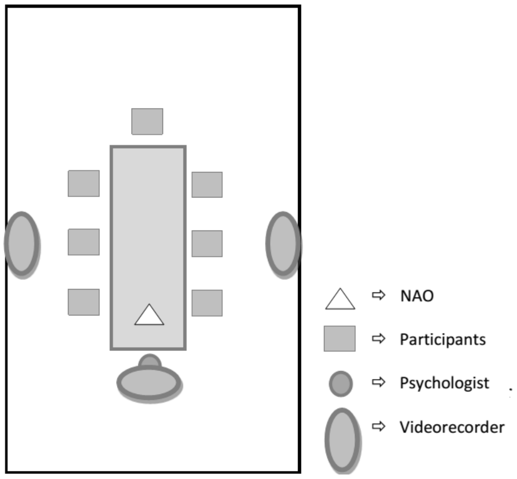

Fig. 3 Layout of the room for the third group.

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);
2. Nao gives a wrong answer (false positives);
3. The patient's answer occurs before Nao is ready to process the words (before the beep);
4. Nao interprets the voice as noise and does not react to the sound (especially if there are several voices that speak simultaneously).

With 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, to capture as much as possible the interaction of participants with the humanoid. In this case, the application was modified to avoid automatically handling the question to be asked; instead, 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

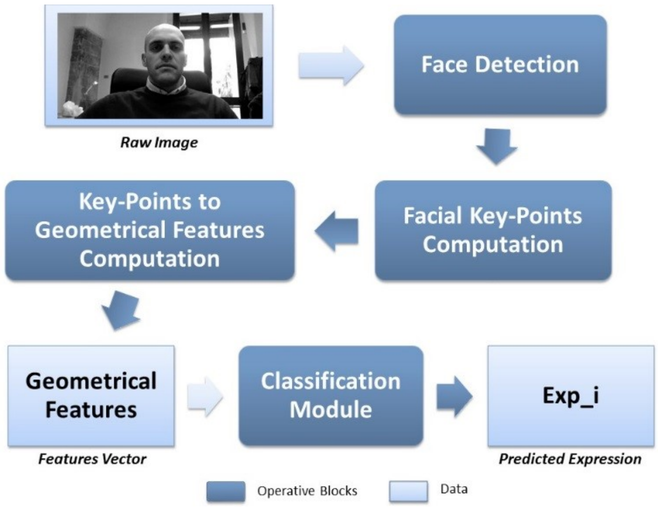

Fig. 4 Software flow diagram.

are more subject to error, since they are open questions and their response can be expressed in multiple ways. 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.

In order 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 Figure 3. The video recordings were cut into fragments of equivalent length and analyzed with a specific software to compare the interaction between the subjects and Nao or the psychologist [29]. This software identifies one or more faces in the video calculating the frequency with which a specific expression appeared (in our case a smile) or turns towards Nao or towards the psychologist. As shown in the flowchart (Figure 4), after identifying an expression, the software defines it and some 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, to the appearance in the video.

Once the 77 points are identified, the software tracks 15 linear, poligonal, eliptical and angular characteristics (Figure 5), 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; 1 for the mouth. Thus, the software traces the

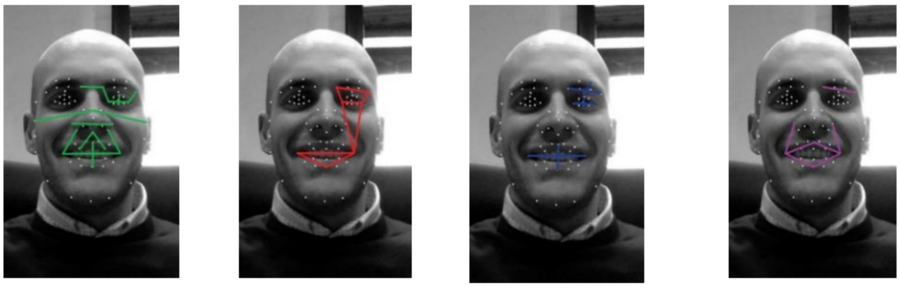

Fig. 5 Linear, polygonal, elliptical and angular characteristics explored by the software in the analysis of human-robot interaction.

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; 3 for the mouth, lower and upper lips. Finally, to determine if the head is tilted, the software combines several points forming lines, (1 to define eyebrows and 6 to define the mouth), and computes the inclination of the lines to determine the head inclination. 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 the statistical method of random forest, analyzes the geometric characteristic vectors to identify the expression or, in our case, to determine if the subjects smiled during the interactions with Nao or with the psychologist.

Additionally, the software allows establishing gaze direction and head position, returning three coordinates in the space and using the characteristic points of the face (such as the tip of the nose) to estimate the gaze direction. Therefore, knowing Nao's or the psychologist's position, through the intersection of two images taken from the same scene (frontal and lateral), it can be assumed whether the patient is looking at Nao or the psychologist. In conclusion, the information obtained through this software during the exercises for the third group includes the parameters:

1. Frequency of visual attention (defined as the number of times each patient of 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 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 psychologist).

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 measured. Subjects in this group were administered three other questionnaires during 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 Results

3.1 Statistical analysis and participants' characteristics

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. The final number of participants was of 21 individuals (10 females and 11 males); their average age was of 73.45 years (SD = 7.71) and the mean education was of 9.90 years, (SD = 4.58). 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.

3.2 Neuropsychological evaluations

In Tables 1-2 the results of each group for all the neuropsychological tests 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 and standard deviations in Figure 6 for Ability, Adaptability and Self-esteem dimensions for PIADS questionnaire. Table 3 shows the frequency of the different types of errors (per group) occurred for Nao or the participants during the interactions on exercises regarding questions on the story. State anxiety (STAI - X) was also measured after the last interaction with Nao, and the general score mean (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 scores from 40 to 50 indicate a state of mild anxiety, and only scores exceeding 60 points are 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) AN.o.VA. mixed models. The statistical analysis revealed a significant difference for prose memory, with respect to the time in which measurements were executed [F (1, 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

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
A	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)
B	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)
C	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)

group and time [F (2, 18) = 4.243, p ¡0.030] appeared for the verbal fluency scores indicating an increase following treatment.

Table 2 MACQ, Fluency (F), Attention (A) mean scores at T0 and T1, and STAI-X at T1 (Standard Deviations in parenthesis)

Group	MACQ T0	MACQ T1	F T0	F T1	A T0	A T1	STAI-X
A	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)
B	25.3 (2.8)	25.6 (1.3)	29.2 (6.5)	34.5 (9.4)	39.2 (9.4)	42.4 (7.3)	30.8 (14.4)
C	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)

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 PI-ADS (see Figure 6), along the Ability dimension, although all groups have evaluated the tool in a positive way, the first group is much dispersed (X=0.8, SD=1.1) and with an average below 1 point, while values of the second group are very compact (X=0.1, SD=0.3) just above zero. The third group 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's 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 Nao's experience. These scores contrasting with those of the third group (X=1.1, SD=1.1), which are 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, with a mean value close to zero (X=0.2, SD=0.4), did not perceive a change in Self-esteem and had 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 found for this scale was 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.

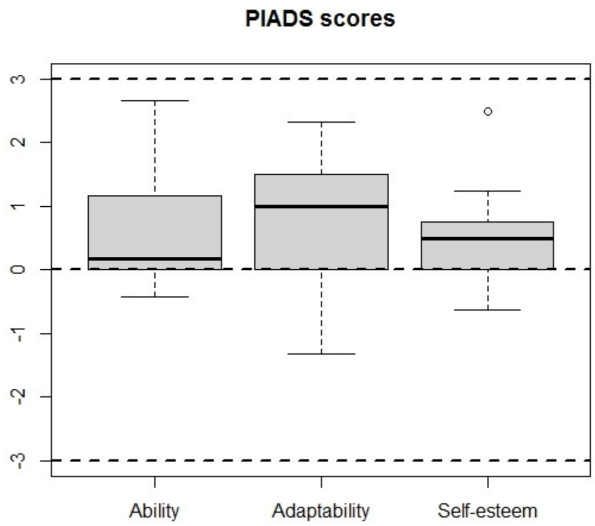

Fig. 6 Mean values for the Psychosocial Impact of Assistive Devices Scales.

3.3 Session monitoring: errors' type

Regarding the exercise 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 the table 3, the first participants' group could not answer all the questions, because once the program cycle was started, it could not be managed, so if Nao comitted 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 subjects' group, 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 group the speech recognition software was inhibited and replaced by the two tactile sensors in Nao's head which, activated by the trainer, threw a positive (if response was correct) or a negative (in case of incorrect response) phrase randomly chosen from two separate lists.

3.4 Participants' experimental performances and human-robot interaction analysis

In Tables 4-5 mean values for frequency and time length of both Smiles and Gazes during the different tasks are reported. The story reading exercise was performed by Nao and the psychologist in their standard position: Nao on

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

the table and the psychologist standing by the narrow part of the table, as displayed in Figures 2-3. In both cases, patients listen, without reading the story. Nao in this exercise did not use its own synthesized voice, but a recorded human voice. As seen in Table 5, no smiles were recorded while the psychologist reads the story, while some smiles were recorded while Nao reads the story. Even the frequency with which patients watch Nao while reading and the lenghth of the gaze is slightly greater, although not significantly. In the patient recognition and naming exercise, the psychologist occupied the typical position around the table, but the Nao's position must necessarily vary to allow him to recognize the patient's facebecause Nao must have the participants 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 on 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 frequencies of the visual interaction towards Nao (M = 35) and the one concentrating towards the psychologist (M = 18.7) [t (12)= 3.99, p = 0.02] and that corcerning the lenght 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 depend 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 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 superiority in 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 bizarre, 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 recall of paired words with the psychologist occurs practically in the same way, with a sheet containing the first word of the paired couples of 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

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

TASKS	NAO Mean (SD)	TRAINER Mean (SD)	p-value Correct
Story Reading			
Frequency	.7 (0.90)	0 (0)	0.38
Time	3.9 (6.20)	0 (0)	0.59
Recognition & 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)	* 0.04
Time	9.70 (11.44)	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

participants, one at a time, state the response. Following the response, 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).

Table 5 Mean values for frequency and time length of gazes 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 & Naming			
Frequency	36 (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

The song-singer matching is an exercise in which participants have to remember the song's title to the singer who made that piece famous. The procedure with the psychologist is performed matching title and singer writing the response. In the procedure for the experimental condition, instead, Nao sings the song with the original singer's voice, and waits for a spoken response (the name of the singer) to process it and deliver the feedback about the 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].

4 Conclusions

In the present study, a humanoid robot was used as support in a memory-training program addressed to individuals with Mild Cognitive Impairment. 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 changes were expected in the neuropsychological tests. It has also been hypothesized that elderly who were supported by Nao during the exercises would experience higher levels of subjective memory self-evaluation and lower levels of anxiety than elderly without such support. Therefore, no significant differences were expected in their reaction during the exercises 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 questionnaire, perhaps due to the heterogeneity of the sample, including individuals affected by all types of MCI and not only the amnesic one. State anxiety measured following a session with Nao, as expected, showed an average value below the mild anxiety threshold. On the other hand, no significant changes were expected in the other measures. Unexpectedly, significant changes were detected in 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's 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 uncontrollable factor. The assistive tool evaluation, in terms of ability to perform actions, adaptability to the environment and self-esteem, produced a positive score. In all the three dimensions, the third group showed a higher average, while the second group scored a lower mean value. The System Usability Scale qualified Nao as an easy-to-use tool. In both cases (PIADS and SUS), patients qualify the instrument's usability as the combination of humanoid and relevant programming, since Nao was in some cases managed directly by the operator bypassing Nao's interactive skills. Errors made by Nao and patients during the two exercises of open questions were taken into consideration, because that type of exercise involves complex interaction. The data shows that inhibiting the speech recognition software, errors disappeared. As expected, no significant differences were found in the patient's behavior when reading a story; in both cases (Nao vs. psychologist), it 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 similar. The significant differences found in the recognizing and naming exercise may be due to the necessary 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 present differences in the number of times that 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 managed by the psychologist is done with paper and pencil; whereas, with Nao, the stimulus was not visual, but acoustic. Also, 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 during sessions, 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 occasions, enthusiastic behaviors were observed on patients as a reaction to some reinforcement phrases after an exercise, rather than during the exercise itself, as long as the reinforcement phrases were not repetitive, but casually chosen from a list of general reinforcements. It would be interesting to personalize these sentences with the name of the patient who answered 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 exercises should be redesigned. In our layout, patients can sit in a fixed place during the sessions, and take advantage of the patient's 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 done cyclically, observing and using feedback from the presentation of an exercise to make it increasingly variable and close to human behavior, until Nao will be able to move independently without the programmer's intervention. Nao's behavior variability is important with the aim to enrich the interaction with patients, so it would be appropriate to initially prepare different versions of the same task, with different movements, different attitudes and different intonation of the voice, for then choose randomly between the programs or between programs' portions, in order to always have a minimum variability in the execution of the same exercise. This could provide the typical freshness of human beings when reading the same sentence twice or supplying the same explanation several times without being exactly the same way as the previous one. Nao's programming it is usually not considered during all the time it does not perform an exercise, for example, when it arrives in the lab. 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 exercise 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 his interactions with the patient in free situations (where the humanoid not performs 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 silenced 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 by an interdisciplinary group, where the work of those who develop and manage the training procedure and study human behavior is complemented by creative people who write and imagine variable scripts for the robot that can be launched in that specific environment.

Author Contributions 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 was in charge of software assessments and analysis. OP and BC were in charge of manuscript verifying.

Conflict of Interest Statement The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

The authors declare that they have no conflict of interest

References

1. A, F., D, G., A, P.: Una palestra per la mente - volume 2 - nuovi esercizi di stimolazione cognitiva per linvecchiamento cerebrale e le demenze (2011)
2. Anzalone, S.M., Tilmont, E., Boucenna, S., Xavier, J., Jouen, A.L., Bodeau, N., Maharatna, K., Chetouani, M., Cohen, D., Group, M.S., et al.: How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot. Research in Autism Spectrum Disorders **8**(7), 814–826 (2014)
3. Bäck, I., Makela, K., Kallio, J.: Robot-guided exercise program for the rehabilitation of older nursing home residents. Annals of Long Term Care **21**(6) (2013)
4. Chang, C.W., Lee, J.H., Chao, P.Y., Wang, C.Y., Chen, G.D.: Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. Journal of Educational Technology & Society **13**(2) (2010)
5. Crook, T.H., Feher, E.P., Larrabee, G.J.: Assessment of memory complaint in age-associated memory impairment: the mac-q. International psychogeriatrics **4**(2), 165–176 (1992)
6. De Carolis, B., Ferilli, S., Palestra, G.: Simulating empathic behavior in a social assistive robot. Multimedia Tools and Applications **76**(4), 5073–5094 (2017)

7. Donata. Ferrari Arianna. Peruzzi, A.G.: Una palestra per la mente. Stimolazione cognitiva per l'invecchiamento cerebrale e le demenze. Edizioni Centro Studi Erickson (2007)
8. Ertuğrul, B.S., Gurpinar, C., Kivrak, H., Kose, H.: Gesture recognition for humanoid assisted interactive sign language tutoring. In: Signal Processing and Communications Applications Conference (SIU), 2013 21st, pp. 1–4. IEEE (2013)
9. Fernandes14, E., Fermé, E., Oliveira, R.: Using robots to learn functions in math class. Technology Revisited p. 152 (2006)
10. Gallese, V., Keysers, C., Rizzolatti, G.: A unifying view of the basis of social cognition. Trends in cognitive sciences **8**(9), 396–403 (2004)
11. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al.: Mild cognitive impairment. The Lancet **367**(9518), 1262–1270 (2006)
12. Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C.: The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. Neuroimage **35**(4), 1674–1684 (2007)
13. Graf, B., Parlitz, C., Hägele, M.: Robotic home assistant care-o-bot® 3 product vision and innovation platform. In: International Conference on Human-Computer Interaction, pp. 312–320. Springer (2009)
14. Gross, H.M., Schroeter, C., Mueller, S., Volkhardt, M., Einhorn, E., Bley, A., Langner, T., Martin, C., Merten, M.: 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. IEEE (2011)
15. Gross, H.M., Schroeter, C., Mueller, S., Volkhardt, M., Einhorn, E., Bley, A., Langner, T., Merten, M., Huijnen, C., van den Heuvel, H., et al.: 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. IEEE (2012)
16. Ismail, L.I., Shamsudin, S., Yussof, H., Hanapiah, F.A., Zahari, N.I.: Robot-based intervention program for autistic children with humanoid robot nao: initial response in stereotyped behavior. Procedia Engineering **41**, 1441–1447 (2012)
17. J, B.: Sus: A quick and dirty usability scale. In: T. B., W. A., M.I. I. (eds.) Jordan P. W., pp. 189–194 (1996)
18. J, J., H, D.: Psychosocial impact of assistive devices scale (PIADS). Technol Disabil. 14. 107-111 **14.**(10.), 107–111 (2002)
19. Kidd, C.D., Breazeal, C.: Robots at home: Understanding long-term human-robot interaction. In: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pp. 3230–3235. IEEE (2008)
20. Kory, J., Breazeal, C.: Storytelling with robots: Learning companions for preschool children's language development. In: Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, pp. 643–648. IEEE (2014)
21. Martín, F., Agüero, C., Cañas, J.M., Abella, G., Benítez, R., Rivero, S., Valenti, M., Martínez-Martín, P.: Robots in therapy for dementia patients. Journal of Physical Agents **7**(1), 48–55 (2013)
22. Montazeri, A., Vahdaninia, M., Ebrahimi, M., Jarvandi, S.: The hospital anxiety and depression scale (hads): translation and validation study of the iranian version. Health and quality of life outcomes **1**(1), 14 (2003)
23. Mubin, O., Stevens, C.J., Shahid, S., Al Mahmud, A., Dong, J.J.: A review of the applicability of robots in education. Journal of Technology in Education and Learning **1**(209-0015), 13 (2013)
24. Novelli, G., Papagno, C., Capitani, E., Laiacona, M., et al.: Tre test clinici di ricerca e produzione lessicale. taratura su sogetti normali. Archivio di psicologia, neurologia e psichiatria (1986)
25. Novelli, P., Capitani, L., Vallar, C., Cappa, S.: Test di fluenza verbale. Archivio di Psicologia, Neurologia e Psichiatria **47**(4) (1986)
26. O, P., R, T.: Tecnologie assistive: Nao robot come interfaccia per la stimolazione cognitiva e lautonomia di anziani con indebolimento cognitivo. 12a Conferenza Nazionale GIMBE: Per una sanit al centro dellagenda politica. Bologna **3** (2017)

27. Olimpia, P.: Ricucire i ricordi: la memoria, i suoi disturbi, le evidenze di efficacia dei trattamenti riabilitativi. Mondadori università (2017)
28. Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., Vallar, G.: Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. The Italian Journal of Neurological Sciences **8**(6), 537–548 (1987)
29. Palestra, G., Pettinicchio, A., Del Coco, M., Carcagnì, P., Leo, M., Distante, C.: Improved performance in facial expression recognition using 32 geometric features. In: International Conference on Image Analysis and Processing, pp. 518–528. Springer (2015)
30. Petersen, R.C.: Mild cognitive impairment: current research and clinical implications. In: Seminars in neurology, vol. 27, pp. 022–031. Copyright© 2007 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. (2007)
31. Pino, O.: Memory impairments and rehabilitation: Evidence-based effects of approaches and training programs. The Open Rehabilitation Journal **8**(1) (2015)
32. Pino, O., Poletti, F., Caffarra, P.: Cognitive demand and reminders effect on time-based prospective memory in amnesic mild cognitive impairment (amci) and in healthy elderly. Open Journal of Medical Psychology **2**(01), 35 (2013)
33. Spielberger, C.D.: State-trait anxiety inventory for adults. PsycTESTS Dataset (1983). DOI 10.1037/t06496-000
34. Spinnler, H.: Standardizzazione e taratura italiana di test neuropsicologici. Ital J Neurol Sci **6**, 21–120 (1987)
35. Ton, T.G., DeLeire, T., May, S.G., Hou, N., Tebeka, M.G., Chen, E., Chodosh, J.: The financial burden and health care utilization patterns associated with amnestic mild cognitive impairment. Alzheimer's & Dementia **13**(3), 217–224 (2017)
36. Triebel, R., Arras, K., Alami, R., Beyer, L., Breuers, S., Chatila, R., Chetouani, M., Cremers, D., Evers, V., Fiore, M., et al.: Spencer: A socially aware service robot for passenger guidance and help in busy airports. In: Field and service robotics, pp. 607–622. Springer (2016)