2019-11

Evaluating the Acceptability of Assistive Robots for Early Detection of Mild Cognitive Impairment

Luperto, M

10.1109/iros40897.2019.8968234
2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
IEEE

2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Macau, China, November 4-8, 2019

Evaluating the Acceptability of Assistive Robots for Early Detection of Mild Cognitive Impairment

Matteo Luperto[1], Marta Romeo[2], Francesca Lunardini[3], Nicola Basilico[1], Carlo Abbate[4], Ray Jones[5], Angelo Cangelosi[2], Simona Ferrante[3], N. Alberto Borghese[1]

Abstract— **The employment of Social Assistive Robots (SARs) for monitoring elderly users represents a valuable gateway for at-home assistance. Their deployment in the house of the users can provide effective opportunities for early detection of Mild Cognitive Impairment (MCI), a condition of increasing impact in our aging society, by means of digitalized cognitive tests. In this work, we present a system where a specific set of cognitive tests is selected, digitalized, and integrated with a robotic assistant, whose task is the guidance and supervision of the users during the completion of such tests. The system is then evaluated by means of an experimental study involving potential future users, in order to assess its acceptability and identify key directions for technical improvements.**

I. Introduction

The development of Socially Assistive Robots (SARs) is a challenge that, in the last few years, has been targeted by many research efforts seeking the goal of having a new, useful, and better-accepted technology for remote patient monitoring and assistance in everyday tasks [1]. Systems based on SARs represent a promising technology aimed at providing assistance to human users through social interactions, and their use has been widely explored and proved to be effective in works such as [2]. The employed robots are often based on autonomous mobile platforms that, thanks to a high level of integration into Ambient Assisted Living environments (AAL), are able to convey services such as the delivery of messages and reminders, teleconferences with family members or caregivers, and guidance through physical or cognitive exercises [3].

A constantly growing number of application scenarios where SARs play a central role find their motivations in the compelling issues posed by the aging of the global population [4], a constantly growing phenomenon with an increasing impact on health care and society in general. Within this context, SARs could be a valuable platform since they allow to deploy their remote health-monitoring functionalities directly in the elders' comfortable home environment, rather than in a more controlled, but also more expensive and often overpopulated setting, as the one of care homes.

When dealing with the effects of aging, cognitive decline is among the most important concerns [5]. Indeed, this process has a strong impact on the life quality of the elderly and can develop into dementia, usually by crossing an intermediate phase, commonly referred to as Mild Cognitive Impairment (MCI). There is currently no effective treatment for dementia, but available therapies have proven to be more efficacious in the early stages of cognitive weakening, showing the ability to extend the duration of independent living [6]. At present, neuropsychological assessment is the most useful tool for identifying patients with MCI. Assessment is usually carried out by means of standardized *cognitive tests*, which take place in the scope of a medical evaluation, i.e., under the supervision of a clinician and in a controlled environment.

While cognitive tests are effectively employed to recognize developed stages of MCI, early assessment of the disorder still represents a challenging task for which an established solution is largely missing. One central reason is that the need for a clinical context to perform the diagnosis causes evaluations to be scheduled only after evident symptoms have already manifested. Conducting cognitive tests in hospitals or clinical facilities has another potential disadvantage, since controlled clinical environments may interfere with the behaviour of the patient, potentially jeopardizing test validity [7]. Administering tests in a familiar home environment can play a fundamental role in overcoming these limitations. To tackle this need, the EU-funded MoveCare project aims at developing a multi-actor system to support the independent living of elderly people through monitoring, assistance, and stimulation of the users in their homes [8].

As part of such a project, the current work investigates whether robotic assistants can provide user guidance during the administration of cognitive tests to elderly users, in their own living environment. More precisely, we devise a system where a robotic platform is integrated with two selected digitalized cognitive tests with proven validity. We show how, thanks to such integration, the robot acquires the ability to undertake real-time interactions with the user by means of simple and structured dialogues. Such interactions are designed to provide guidance through the execution of the cognitive tests, in the attempt of preserving the validity of cognitive assessment without a real clinician. It is important to point out that our system is not a replacement of clinical cognitive assessments nor is actually aiming at performing a cognitive assessment. It is rather an instrument that, if used at home, could ease the first step towards a proper clinical evaluation, thus increasing the chances for early MCI diagnosis.

Our system's evaluations mainly focus on the *acceptability*

[1] Dep. of Computer Science, University of Milan
[2] School of Computer Science, The University of Manchester and AIST Artificial Intelligence Research Center, Tokyo Waterfront
[3] Dep. of Electronics, Information and Bioengineering, Politecnico di Milano
[4] Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan
[5] School of Nursing and Midwifery, University of Plymouth

978-1-7281-4004-9/19/$31.00 ©2019 IEEE

by potential future users. Such an aspect represents one of the key challenges of elderly-oriented modern technologies, and it is crucial, as in our case, when the system is envisaged with the long term goal of being installed in the elder's house and providing several functionalities [9]. Inspired by this vision, we study the feedback obtained by interviewing a group of selected volunteers living in assisted facilities who tested our system, also conducting a comparison with a traditional setting where tests guidance is provided by a clinician. Our results show how participants perceive the opportunity of being supported by a robot during neuropsychological tests at home and, at the same time, allowed us to identify key directions of development in the pursuit of a fully deployable system. A preliminary version of our system has been presented as a live demonstration in [10], where a multimodal interface has been used in place of the robot assistant.

II. Related work

An exhaustive review of previous work exploring the benefits of SARs in elderly care can be found in [11], where a functional distinction is outlined between *service* robots, aiming at helping users in daily life activities, and *companion* robots, targeting the psychological well-being of their owners.

The review highlights a trend that leverages both service and companion robots for mental health care interventions [12] within the residential living environment. The first category includes work like [13], which proposed the use of a half-bust robot to assist the cognitively-impaired elder during mealtime, or [14], in which an info-terminal robot was used to provide useful information and reminders to the residents of a care home. On the other hand, a well-known example of a companion robot is the seal PARO [15], which was primarily used to ease distress in elders suffering from mild or severe cognitive impairments. In this framework, most of the proposed solutions have proved effective in enhancing the well-being of the elder users interacting with the robotic platforms.

Despite the established benefits of using SARs in the context of residential living, the ultimate goal of a new model of preventive care should be the deployment of robotic assistants to the users home for remote health monitoring functionalities [16], [17]. To answer this need, the integration of robots in AAL environments has been proposed in work such as [18], where the teleoperated robot Giraff was deployed to the elder's home, together with a network of sensors, to achieve monitoring of daily-life activities.

In this context, the use of SARs for home-based monitoring of cognitive functionalities with the final aim of detecting early alerts represents an interesting and important application. However, the validity of such a robotic psychometric approach is still an open challenge. For example, in [19], authors show how these robots can provide advantages for early MCI detection thanks to their capabilities of administering specific standardized tests and automatically recording the answers for further analysis while engaging the user. The study stems from an experimental setting where a number of healthy adults completed the "Montreal Cognitive Assessment" (MoCA) [20] under the guide of the humanoid robot Pepper. The limits of [19] includes the fact that the system evaluation was done by healthy adults; in addition, test administration by the robot was standardized and regulated by internal timers, so that it could not be adapted according to the users reaction.

A further step toward the use of SARs for early MCI detection is presented here. Our work aims at developing and testing, on potential elder users, a scenario where Giraff, a versatile assistive robotic platform, assists older adults through the execution of digitalized cognitive tests. Giraff's guidance is achieved through real-time and autonomous interaction with the users, that faithfully mimics the instructions provided during the correspondent clinical protocol.

III. Proposed System

To evaluate how assistive robots can play a role in at-home cognitive assessment, we base our system upon the implementation and integration of the two key elements of the clinical practice: the administration of cognitive tests and its real-time supervision by a caregiver. We provide the first by developing a tablet-based setup for the digital counterparts of standard clinical tests, while the real-time guidance is carried out by an assistive mobile robot: the Giraff platform. This separate deployment of the two functionalities reflects our long-term vision, where the supervision of cognitive assessment is meant to enrich a wider pool of social behaviors that the robot undertakes for its assistive tasks. In such a setting, the robot embodies and translates into interactions the intelligence of the system, while the tablet offers an input interface. This view is also strengthened by works like [21] stating that users respond more positively to robots than tablet computers delivering health care instructions.

A. Cognitive assessment with digitalized tests

Clinical cognitive assessment is customarily supported by *cognitive tests* where the subject is asked to carry out a particular task under the watch of a caregiver [22]. The tests follow standard and validated protocols that rule the supervising interventions, and that define the metrics to consider in the evaluation phase. The tasks can be various (based on paper and pencil, requiring verbal interactions, etc.) and the metrics can be quantitative and qualitative, the latter being more subject to the caregiver's judgment.

The first step of our work focused on the selection of suitable cognitive tests that could be digitalized and included in our reference scenario. We considered the related clinical state of the art on this subject with the support of clinical professionals in the cognitive assessment of elder patients from the Ospedale Maggiore Policlinico hospital. The technical desiderata that guided our choices, and that were discussed with the experts, are the following:

- the digital tests should be easy to interact with, and conveyed by standard consumer technology (like tablets),
- the caregiver's supervision should be structured by simple actions and triggered by clearly-recognizable events,

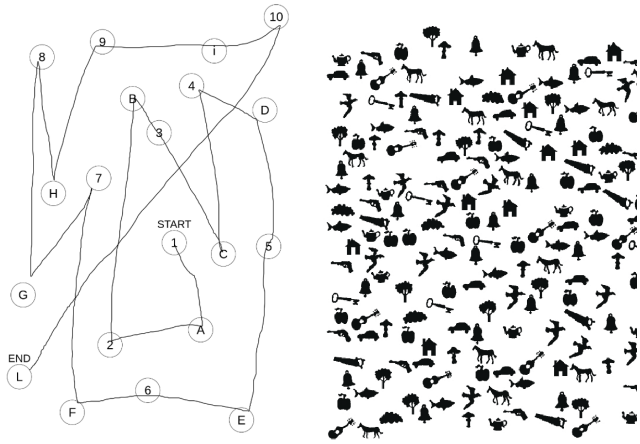
- the evaluation metrics should be primarily quantitative and allow automated computation.

Guided by the above criteria, our attention narrowed down to paper-and-pencil tests since the digital implementation of this input modality can take advantage of many robust development technologies. Among those, we privileged tests where the caregiver's interventions could only minimally impact on the test's evaluation metrics to allow for an easier interpretation of the results. Our final choices have been the Trail Making Tests (TMT) [23] and the "Bells" Test (BT) [24]. Their input interface is provided by a 10-inches tablet endowed with a capacitive pen.

The TMT requires to draw a continuous line traversing a number of symbols according to the sequence suggested by their labels e.g., $1, A, 2, B, \ldots, 10, L$. Directions on how to execute the test are provided with a simple tutorial, where a limited number of symbols are displayed, and where the first symbols are connected by the practitioner as indications. The evaluation metrics should indicate whether the line has been correctly drawn, and the amount of time required to complete the test. Errors made during the execution of the test are signaled, but not corrected. Figure 1a depicts an example of the execution of the TMT from our digital version. The test layout is similar to the original paper-based TMT but, given the reduced dimensions of the tablet screen (the original test is performed on an A4 paper), the number of targets was decreased to 20 as proposed by [25].

The "Bells" test requires to localize, by drawing a mark and within a time budget, a set of identical icons shaped like a bell and mixed up among similar icons called "distractors". As in the case of TMT, a simple tutorial with a limited number of icons is provided, where the test protocol is explained. Errors made during the execution of the test are not corrected nor signaled. The test is considered completed if the user believes that all of the bells have been identified. The evaluation metrics include the number of bells correctly marked, the number of errors, the distribution of the bells that have been found/missed, and the time required to complete the test. The layout of our digital implementation, whose pattern is shown in Fig. 1b, places 35 targets and 280 distractors as done in [24]. Also in this case, the graphical elements have been scaled to fit the tablet's screen size.

TMT and BT are tests typically included in batteries for MCI detection and have been proven useful for assessing early stages of it even on healthy subjects [26]. This is particularly true for the TMT, of which several digital versions have been proposed [27]. These implementations, however, are mere transpositions of the original paper-and-pencil tests over a digital device. The context in which they are meant to be administered is the same as the original tests: in an unfamiliar environment and in the presence of a trained caregiver. The digital versions of TMT and BT that we propose in this work support real-time integration in an assistive robotic scenario where the role of the caregiver is filled by a robot assistant. Besides this interaction, the test digitization allows for automated computation of the test evaluation metrics. Additional details on the implementations



(a) TMT execution (b) BT pattern, from [24]

Fig. 1: The digital cognitive tests of the proposed system.

of our digital tests can be found in [28] where their validity (providing predicting capabilities comparable to those of their traditional counterparts) is shown.

B. Integration with the robot caregiver

Giraff [29], the robot we adopted, was originally designed as a teleoperated platform to provide assistance in domestic environments. As introduced above, our setup exploits it to provide supervision during the execution of TMT and BT (see Fig. 4). We build upon a modified ROS-based [30] version of Giraff developed within the scope of the MoveCare project where new functionalities (like navigation) enable its use as a fully-autonomous assistive robot [9].

Our objective is to equip the robot with a fully autonomous supervising behavior allowing it to explain the test's protocol, provide interventions during the test (as dictated by the test's protocol), and give a final feedback to the user. This objective poses the need for a real-time integration between the digitalized tests and the robot that we achieve with the architecture proposed in Fig. 2.

In such an architecture, cognitive tests are implemented as a Model-View-Controller web app served to a browser-based client running on the tablet (Test App). The Tests Server provides the app's view via HTTP and, by means of a WebSocket channel, keeps track of the current state of the test execution. User's inputs collected by the tablet and all the test's relevant events (start drawing, stop drawing, errors, etc.) are transmitted in real-time to the Tests Server where they are processed and stored.

On the other side, Giraff has been equipped with a speech interface delivering instructions in an understandable way, recognizing the user's answers and requests, and producing the proper audio responses. The interface is composed of two functional modules: the Speech Module and the Dialog Manager (DM).

The Speech Module operates at signal-level providing real-time speech recognition and synthesis. The first exploits

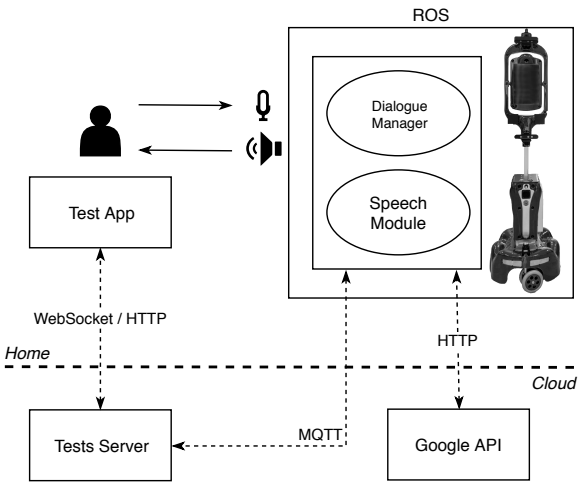


Fig. 2: System architecture.

Google's Cloud Speech-to-Text API[1] to convert into text the user's utterances recorded through the robot's microphone. Subsequently, the DM compares the obtained text with a pre-compiled set of relevant keywords to identify the proper dialogue transition. For example, the "yes" and "understood" keywords correspond to the same semantic transition triggered by a positive answer to the question "Did you understand?". Speech synthesis is based on Acapela Voice As a Service installed as a local service on the robot[2]. This module also exploits the Giraff's 13.3 inches portrait screen, originally designed for telepresence. Indeed, the screen of the robot is used (in parallel with the speech) to display pop-up messages with the most important notifications needed during the execution of the tests. This choice was made in the attempt of compensating the problems that people with hearing impairments may have while listening to the robot's spoken messages.

The DM encloses the high-level dialog logic, guiding the robot's speech interactions. It allows the robot to determine when to listen (opening the mic) and how to answer (through the speakers). It is based on a finite state machine that tracks the current state of the test execution and of the dialog engaged with the user. From one side, the DM receives updates on the test execution by the Tests Server and keeps this last one aware of the current state of the dialog. The integration between the Tests Server and the ROS node executing the DM is obtained by a bi-directional publish/subscribe channel based on the MQTT protocol[3]. The keywords extracted by the speech module provide a second input to the DM that, by considering them, can translate the user's utterances into transitions of the dialog's current state. By accomplishing these tasks, the DM can trigger the correct spoken responses to be delivered by the speech module.

Our system's architecture allows the test app provided by the tablet and the robot's speech interface to work in real-time cooperation. The result is the integration between the digitalized tests and the robot, allowing this last to undertake a supervising behavior dictated by these protocol steps:

1) asking the participant whether he/she wants to proceed with the test;
2) giving the instructions to complete the test;
3) asking the participant if everything was clear;
4) if yes, asking for a confirmation to proceed with the actual test; if not, explaining the test again;
5) during the TMT, informing the participant whenever a mistake is made (during BT errors are not signaled);
6) if the user is idle and not proceeding with the test, asking whether to continue or leave;
7) once the Tests Servers signals that the test is completed (or if the user says something semantically equivalent to "I have finished", during BT only), giving a non-informative feedback, not revealing the final score to prevent distress.

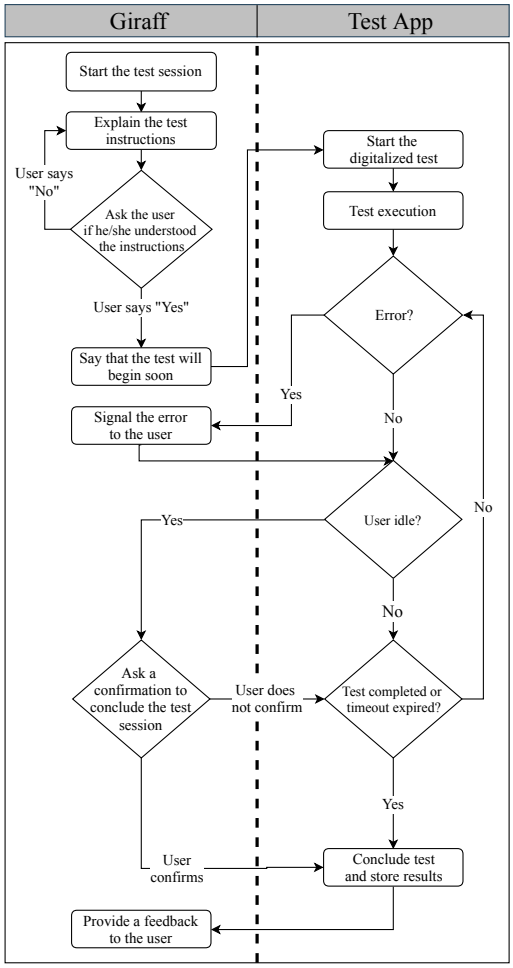


Fig. 3: Interaction between Giraff and the Test App for TMT.

Figure 3 provides the complete picture of how digitalized tests and the robot interact. (The example is related to the TMT). No additional behaviors were required for the robot, which must remain silent for the duration of the tests, unless an error or an idle situation is detected. For this particular

[1]https://cloud.google.com/speech-to-text/
[2]http://www.acapela-vaas.com/
[3]http://mqtt.org/

application, the robot does not need social behaviors and it cannot give too many feedback to the participants while they are completing the task. Introducing biases, distractions, or guiding too much towards a successful completion of the tests are outcomes that need to be prevented.

IV. Experimental evaluation

Since usability is one of the key factors influencing acceptability, we decided to evaluate our system both in terms of acceptability and usability. To do so, we devised an exploratory study involving potential future users. In addition, to distinguish the contribution of the two building blocks (digital tests and robotic supervision) on the obtained scores, the results were compared to the ones collected from a control group, in which potential users performed the digital tests in a clinical environment, under the supervision of an expert human operator [28]. In the following, we first detail our experimental setup and then we outline the results and lessons learned.

A. Experimental session

The experiment used to evaluate the system comprises two different stages: the tests completion and a follow-up interview. The protocol was approved by the Ethical Board of Plymouth University.

Participants were recruited on a voluntary basis from the elderly people living in apartments that are part of a shelter house that provides assisted living in Plymouth. In this kind of facilities, the residents are still completely independent, living alone or with their partner in private apartments, but they can always rely on the presence of a manager, supervising the house and making sure everyone is safely monitored. The only inclusion criteria adopted was old age (above 65) and, given the "qualitative" nature of the answers collected through interviews, no constraints on the sample size were imposed [31]. In total, 16 participants took part in the experiment (4 males and 12 females). The age range was between 65 and 86 (*Mean=73.6*, *Standard Deviation=5.9*). The system setup was the one shown in Fig. 2, where cloud connectivity was provided by a portable 4G hotspot since no Wi-Fi or Ethernet connections were available in the shelter housing where the experiment took place.

After signing a consent form, participants were asked to take part in the experiment, with Giraff as their guide. The investigator was present in the room for the entire duration of the session, to explain the aim of the study and how the experiment would have been carried out, making sure that the platform was working correctly, and answering any questions from the participants. In particular, it was made clear that the aim of the experiment was assessing the acceptability and the usability of the Giraff robotic platform as a guide for the completion of cognitive tests. Thus, the performance of the participants on the tests was not monitored, no indicator computed and no records on the performance saved. Once the participants had clearly in mind that the main aim was not monitoring them but evaluating the system, they could start interacting with the platform and move to the first stage of the evaluation.

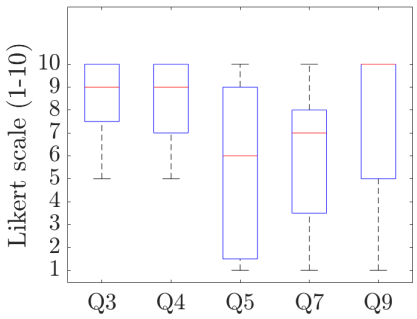
Fig. 4: A participant completing the tests.

During this phase, each participant was asked by the robot to complete TMT and BT on the tablet (Fig. 4) following the instructions given by the robot as dictated by the protocol's steps. The role of the investigator during this stage was limited to making sure that participants were always at ease in continuing the experiment.

After the completion of the last cognitive test, the second stage started. In this phase, the investigator interviewed the participants on their experience with the whole system, through a questionnaire composed of 10 questions, five of them open and five Likert-based:

(Q1) How familiar are you with technology?
(Q2) How familiar are you with robots?
(Q3) On a scale from 1 to 10, 1 being extremely difficult and 10 being perfectly fine, how easy was to understand the instructions given by the robot?
(Q4) On a scale from 1 to 10, 1 being extremely difficult and 10 being perfectly fine, how easy was to use the tablet computer interface?
(Q5) On a scale from 1 to 10, 1 being absolutely frustrating and 10 being perfectly pleasant, how would you rate your experience today?
(Q6) What do you consider the most frustrating factor of todays experience?
(Q7) On a scale from 1 to 10, 1 being absolutely not and 10 being absolutely yes, how much would you consider having a robot as an additional monitoring tool, suggesting and guiding you through different kind of tests?
(Q8) Why?
(Q9) On a scale from 1 to 10, 1 being completely uncomfortable and 10 being perfectly fine, if you were given a robot like the one you used today to keep in your house, and you knew that it had the possibility to monitor your cognitive abilities by the use of such tests. How comfortable would you be in accepting its suggestion and start a test?
(Q10) Why?

Apart from (Q1) and (Q2), the rest of the open questions



Question	Mean	SD
(Q3)	8.625	1.5
(Q4)	8.25	1.8
(Q5)	5.5	3.6
(Q7)	6	3
(Q9)	7.5	3.5

Fig. 5: Score distributions.

were designed to ask for more detailed explanations of Likert scores. The method chosen to record the answers of the participants, an interview with the investigator who assisted the users in mapping their judgments to the scores, and the use of open questions together with the Likert, were both means to allow respondents to elaborate and qualify their answers, giving more information on the reasons behind their choices. In particular, four questions assessed the usability of the system (Q3)–(Q6). The remaining ones (Q7)–(Q10) surveyed the general acceptability of robots, and of the specific system that they had just experienced.

B. Results and Lesson Learned

The important lessons learned for the developers are to be found in the results on the usability of the platform. The average amount of time for the participants to complete the tests was 20 minutes. The investigator had to intervene during the testing phase only six times, three of those to actually explain better to the participants what they had to do, and the remaining times to supply technical assistance due to a slow connection. Overall, no additional explanation by the investigator was needed for 82% of the participants, meaning that the robot guide was clear enough for them to complete the task. Indeed, as the quantitative data from the Likert questions show in Fig. 5, when asked whether the instructions given by the robot were clear and easy to follow (Q3), 94% of participants answered positively, giving a score grater than five (between 7 and 10) to the intelligibility of the robot. On the same note, when asked to rate the ease of using the tablet and the smart pen (Q4) the 88% declared that they had no problems in completing the tasks with such tools, giving a score between 7 and 10.

(Q4)'s main aim was to assess the usability of the Test App and, specifically, of its GUI, used by participants to complete the test. It was not questioning the hardware, given that both the tablet computer and the capacitive pen used for the experiment are commercially available products not developed by the authors. Nonetheless, interesting comments on the usability of the hardware emerged from the interview as well: in particular, the totality of participants had never used a capacitive pen before and two participants explicitly stated that the pen was difficult to use, as they could not find the right angle to use it or they were exerting too much pressure while using it. Although only two participants explicitly reported the problem, it was noted that five participants experienced the same problem and this convinced the investigator that an explanation on how to use the pen would be useful to add in the future.

As far as the graphical interface is concerned, the main comment gathered on its usability was that the BT icons were too small, and this made the test more complicated for six participants (37%), two of whom also stated they suffer from visual impairments. This indication suggests that test layout adaptations (for example, having a less dense set of larger icons for BT) or setup upgrades (for example, introducing a tablet with a bigger screen) might improve the system's usability.

To better interpret the results in terms of acceptability, it is key to highlight that the people taking part in this experiment declared themselves mostly not confident with technology (57%) or comparable to a medium-level user (37%), who is used to deal with computers only to stay in touch with family members (Q1). The majority (82%) had never seen a robot before, or just saw some examples on the media (Q2). Despite this, it is interesting to notice that, after using the platform, the 75% of the participants declared they would keep a robot in their houses (Q7), if not in the immediate future, at least when they would start to realize that they need an additional monitoring system to take care of them.

In addition to this, from the comments gathered through (Q9) and (Q10), it emerges that only the 18% of the participants wouldn't feel comfortable in doing such tests with a robot that was actually measuring their abilities. The main reason given by them for not accepting the real monitoring was stated to be the violation of their privacy. Therefore, they would not trust a machine actively assessing their abilities. Despite this, it is a positive result for this research to see that even people with little or no familiarity with robots and technology, after trying the system out, were convinced of the utility of it. In particular, of the thirteen people that declared to have no familiarity at all with robots, nine (70%) said they would keep a robot in their houses, and eleven (85%) stated they would feel comfortable in taking the cognitive tests with the robot. Importantly, comments like "It is useful to have the possibility to do this at home and not to move to go to the hospital" or "It is a robot, it doesn't make any difference" were reported by such users.

TABLE I: Distribution of the scores for (Q5), (Q7), (Q9).

Q	1	2	3	4	5	6	7	8	9	10
Q5	4	1	1	1	1	0	1	1	5	1
Q7	2	2	0	0	2	0	3	5	1	1
Q9	2	0	1	0	3	0	0	0	0	10

The answers gathered from the open questions, especially from the participants who reported a lower acceptability, were key to gain insights into the obtained results including the high variability of the answers to questions like (Q5),

TABLE II: Results from Robot-assisted and Clinician-assisted performance of the digitized cognitive tests

	Q1			Q3		Q4		Q5	
	Low [%]	**Medium [%]**	**High [%]**	**Positive (>5) [%]**	**Mean ± SD**	**Positive (>5) [%]**	**Mean ± SD**	**Positive (>5) [%]**	**Mean ± SD**
Robot-assisted	0.56	0.38	0.06	0.94	8.6 ± 1.5	0.88	8.3 ± 1.8	0.50	5.5 ± 3.6
Clinician-assisted	0.33	0.62	0.05	1.00	9.3 ± 1.1	1.00	8.7 ± 1.3	1.00	8.8 ± 1.4

(Q7), and (Q9). For example, looking at the distribution of the scores for (Q5) in Table I, seven participants (50%) rated negatively (giving a score lower than 5, value used to express neutral opinions) the experience they had while taking part in our experiments. When analyzing the reasons why they gave this evaluation, it was clear that four of them (50% of those who gave bad scores) encountered some difficulties with the speech recognition module of the robot, failing due to the slow connectivity that sporadically affected our setup. A stable Internet connection is an essential requirement, and shortages of such a service can rapidly turn a smooth and entertaining interaction into a more tedious one. This is also confirmed by the fact that four of the participants giving a negative score to (Q5) have then given a score ranging from 5 to 10 to questions (Q7) and (Q9) stating, for example, that "It's good to have the possibility of having robots in the house, they can be useful". Despite these positive scores for (Q7) and (Q9), also the set of answers to such questions were rather polarized. Examples of explanations over the decision to give a negative or neutral score were: "I do not need a robot now, maybe in the future" or "The robot is too big for my house, if it was smaller it would be better". Finally, only two participants were truthful to their negative scoring throughout the questionnaire, clearly showing no willingness to accept the system and the robot, with comments like "I am used to my way of doing things, I like to go out, I will never accommodate to something that bosses me".

After having analyzed the comments for each question, it became clearer that, overall, the system proposed was accepted by the majority of the participants taking part in the experiment. Statement supported also by the fact that ten participants, the 62% of the total, answered with a score higher than 5 to (Q7) and (Q9) while only four and three participants respectively gave a score lower than 5. Therefore, the most important insight that we can draw from this evaluation study is that there is value in developing systems like the one we propose because users understand its potential benefits and lean towards accepting it.

Such belief is also supported by the comparison between the answers (to (Q1), (Q3), (Q4), and (Q5)) obtained in the current study (Robot-assisted) with the results collected from a previous work of our group [28], in which the same digitized tests were performed at the Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico of Milan by 21 cognitively-healthy older adults (age range: 69-84, *Mean=76.3*, *Standard Deviation=4.3*), under the supervision of a clinician (Clinician-assisted) (Table II). Since the two groups were tested on the same Test App, and given that they were characterized by comparable age range and technology acquaintance, the extremely similar results obtained to (Q4) are not surprising. On the other hand, what is more relevant to the purpose of the current study is the similarity of results emerged for (Q3). Indeed, the test instructions provided were clearly understood by elder participants whether they were conveyed by a human operator, as is our previous study, or by a robotic assistant, as in the current work. Such an important result supports the usability of Assistive Robots in assisting older adults during the performance of cognitive tests in a non-clinical environment. Finally, the between-group difference emerged in terms of overall satisfaction can be explained in light of the comments gathered from the open questions. Indeed, the very low score to (Q5) provided by four participants was probably due to slow speech recognition due to the sporadic connection problems. However, the positive feedback from the same subjects to (Q7) and (Q9) supports the hypothesis that the overall satisfaction score would have benefited from more stable connectivity.

Despite these promising results, there are still several technical difficulties that need to be tackled in further developments of the system. Participants were eager to give more precise feedback and they spotted some technical problems that will be addressed in the future versions of this work. The main issues participants had while completing the tests were mainly due to the repetitive question and answer loops they had to engage with the robot. In order to go through all the tests, the robot was programmed to ask at least twice per test if the participants had understood what they had to do and if they were ready to start the test. This, together with the fact that the speech recognition sometimes failed, led to an additional request by the robot to repeat their answer. This fact made the human-robot interaction slow and repetitive at times, as pointed out by the 50% of participants. One of the main goals of this platform is to keep its users interested and engaged so that the next time the robot proposes the tests they would gladly accept to repeat them. This is the main reason why understanding that the number of interactions needs to be carefully balanced is very valuable information. In future developments, the dialogues will be simplified following the guidelines emerged from our analysis.

In addition, although reduced sample sizes are particularly prevalent in telehealth reports from pilot or feasibility studies [32], in the future collaborative research efforts will be adopted to jointly collect answers from multiple sites.

V. Conclusion

The work presented in this paper aims at developing a non-intrusive system, to be deployed in the house of the elderly,

that helps the detection of early signs of MCI. It consists of a robotic platform able to guide its users through the completion of cognitive tests. We developed and evaluated a prototype system with a group of possible future users in terms of acceptability and usability. Results show how older people understand the importance of additional monitoring, and accept the guidance of a robot, feeling comfortable with it explaining and supervising the tests instead of a clinician.

The promising results we observed in this work will be further investigated for an extended period of time within the MoveCare project's pilot. Among the future working directions we envisage the development of enhanced empathic behaviours for the robot's UI and the comparison against a setting where supervision is carried out by a virtual assistant embedded in the tests' interface.

ACKNOWLEDGMENTS

Authors would like to thank Plymouth Community Homes, and in particular Hazel Alexander, for their help in recruiting care facilities and participants. This work is partially supported by the European H2020 project MoveCare grant ICT-26-2016b GA 732158.

REFERENCES

[1] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *Proc. ICORR*, 2005, pp. 465–468.

[2] A. Tapus, C. Tapus, and M. J. Mataric, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *Proc. ICORR*, 2009, pp. 924–929.

[3] N. Bellotto, M. Fernandez-Carmona, and S. Cosar, "Enrichme integration of ambient intelligence and robotics for aal," in *Proc. AAAI Spring Symposium Series*, 2017.

[4] U. Nations, "World population ageing: 1950-2050," *New York: Department of Economic and Social Affairs*, 2002.

[5] M. M. Johansson, J. Marcusson, and E. Wressle, "Cognitive impairment and its consequences in everyday life: experiences of people with mild cognitive impairment or mild dementia and their relatives," *INT PSYCHOGERIATR*, vol. 27, no. 6, pp. 949–958, 2015.

[6] D. Budd, L. C. Burns, Z. Guo, G. LItalien, and P. Lapuerta, "Impact of early intervention and disease modification in patients with predementia alzheimers disease: a markov model simulation," *Clinicoecon Outcomes Res*, vol. 3, p. 189, 2011.

[7] N. Chaytor and M. Schmitter-Edgecombe, "The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills," *Neuropsychol Rev*, vol. 13, no. 4, pp. 181–197, 2003.

[8] F. Lunardini, N. Basilico, E. Ambrosini, J. Essenziale, R. Mainetti, A. Pedrocchi, K. Daniele, M. Marcucci, D. Mari, S. Ferrante, and N. A. Borghese, "Exergaming for balance training, transparent monitoring, and social inclusion of community-dwelling elderly," in *Proc. RTSI*, 2017, pp. 1–5.

[9] M. Luperto, J. Monroy, F.-A. Moreno, J. Ruiz-Sarmiento, N. Basilico, J. G. Jimenez, and N. Borghese, "A multi-actor framework centered around an assistive mobile robot for elderly people living alone," in *IROS Workshop on Robots for Assisted Living*, 2018.

[10] M. Luperto, M. Romeo, F. Lunardini, N. Basilico, R. Jones, A. Cangelosi, S. Ferrante, and N. A. Borghese, "Digitalized cognitive assessment mediated by a virtual caregiver," in *Proc. IJCAI*, 2018, pp. 5841–5843.

[11] J. Abdi, A. Al-Hindawi, T. Ng, and M. P. Vizcaychipi, "Scoping review on the use of socially assistive robot technology in elderly care," *BMJ OPEN*, vol. 8, no. 2, 2018.

[12] S. M. Rabbitt, A. E. Kazdin, and B. Scassellati, "Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use," *Clin. Psychol. Rev.*, vol. 35, pp. 35–46, 2015.

[13] D. McColl, W. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 74–83, 2013.

[14] M. Hanheide, D. Hebesberger, and T. Krajnik, "The when, where, and how: an adaptive robotic info-terminal for care home residents a long-term study." in *Proc. HRI*, 2017.

[15] K. Wada, T. Shibata, T. Saito, and K. Tanie, "Effects of robot assisted activity to elderly people who stay at a health service facility for the aged," in *Proc. IROS*, 2003, pp. 2847–2852.

[16] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, and M. Vincze, "Hobbit, a care robot supporting independent living at home: First prototype and lessons learned," *ROBOT AUTON SYST*, vol. 75, pp. 60–78, 2016.

[17] H.-M. Gross, S. Mueller, C. Schroeter, M. Volkhardt, A. Scheidig, K. Debes, K. Richter, and N. Doering, "Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments," in *Proc. IROS*, 2015, pp. 5992–5999.

[18] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, C. Galindo, J. Gonzalez, L. Karlsson, A. Forsberg, S. Frennert, F. Furfari, A. Loutfi, A. Orlandini, F. Palumbo, F. Pecora, S. von Rump, A. Štimec, J. Ullberg, and B. Ötslund, "Giraffplus: A system for monitoring activities and physiological parameters and promoting social interaction for elderly," in *Human-Computer Systems Interaction: Backgrounds and Applications 3*, 2014, pp. 261–271.

[19] S. Varrasi, S. Di Nuovo, D. Conti, and A. Di Nuovo, "Social robots as psychometric tools for cognitive assessment: A pilot test," in *Human Friendly Robotics*, 2019, pp. 99–112.

[20] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *J AM GERIATR SOC*, vol. 53, no. 4, pp. 695–699, 2005.

[21] J. A. Mann, B. A. MacDonald, I.-H. Kuo, X. Li, and E. Broadbent, "People respond better to robots than computer tablets delivering healthcare instructions," *Comput. Hum. Behav.*, vol. 43, pp. 112–117, 2015.

[22] E. Strauss, E. M. Sherman, and O. Spreen, *A compendium of neuropsychological tests: Administration, norms, and commentary*. American Chemical Society, 2006.

[23] R. M. Reitan, "Validity of the trail making test as an indicator of organic brain damage," *PERCEPT MOT SKILLS*, vol. 8, no. 3, pp. 271–276, 1958.

[24] L. Gauthier, F. Dehaut, and Y. Joanette, "The bells test: a quantitative and qualitative test for visual neglect." *J. Clin. Neuropsychol.*, vol. 11, no. 2, pp. 49–54, 1989.

[25] A. R. Giovagnoli, M. Del Pesce, S. Mascheroni, M. Simoncelli, M. Laiacona, and E. Capitani, "Trail making test: normative values from 287 normal adult controls," *Neurol. Sci.*, vol. 17, no. 4, pp. 305–309, 1996.

[26] K. L. Votruba, C. Persad, and B. Giordani, "Cognitive deficits in healthy elderly population with normal scores on the mini-mental state examination," *J. Geriatr. Psychiatry Neurol.*, vol. 29, no. 3, pp. 126–132, 2016.

[27] R. P. Fellows, J. Dahmen, D. Cook, and M. Schmitter-Edgecombe, "Multicomponent analysis of a digital trail making test," *Clin Neuropsychol*, vol. 31, no. 1, pp. 154–167, 2017.

[28] F. Lunardini, M. Luperto, N. Basilico, S. Ferrante, K. Daniele, C. Abbate, S. Damanti, D. Mari, M. Cesari, and N. Borghese, "Validity of digital trail making test and bells test in elderlies," in *Proc. BHI*, 2019.

[29] "Giraff: An advanced teleoperation robot." [Online]. Available: http://www.giraff.org/

[30] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, 2009.

[31] K. Kelley, B. Clark, V. Brown, and J. Sitzia, "Good practice in the conduct and reporting of survey research," *INT J QUAL HEALTH C*, vol. 15, no. 3, pp. 261–266, 2003.

[32] D. Langbecker, L. J. Caffery, N. Gillespie, and A. C. Smith, "Using survey methods in telehealth research: A practical guide," *J Telemed Telecare*, vol. 23, no. 9, pp. 770–779, 2017.