Osaka City Med. J. Vol. 66, 13-21, 2020

Can a General-purpose Interactive Robot Detect Poor Cognitive Function? A Pilot Study

Jumpei Maruta[1,2], Yasunori Matsuda[3], Kentaro Uchida[1], Tadashi Ohtomo[4], Hideo Kurozumi[1], Satoshi Nogi[1], Satoshi Akada[1], Aki Nakanishi[2], Miki Shinoda[5], Masatsugu Shiba[6], and Koki Inoue[1,6]

Departments of Neuropsychiatry[1] and Center for Brain Science[6], Osaka City University Graduate School of Medicine; Medical Center for Dementia[2], Osaka City Kosaiin Hospital; Tsuruhashi Mental Clinic[3]; Shigisan Hospital[4]; and Department of General Welfare and Clinical Psychology[5], Osaka City University Graduate School of Human Life Science

Abstract

Background

Neuropsychological examinations have been routinely used to screen for dementia. However, these require human resources, and the subject might feel psychologically burdened by the human examiner. We investigated whether a robot could be used to screen for dementia.

Methods

We used a general-purpose humanoid interactive robot to conduct the Hasegawa Dementia Scale Revised (HDS-R), modified for the robot (robot test). We recruited 12 patients with mild to moderate Alzheimer's disease (AD) and 4 patients with mild cognitive impairment (MCI) from the outpatient department of Osaka City Kosaiin Hospital. We recruited 10 healthy controls (HC) from patients' spouses. The subjects took the robot test, the Mini Mental State Examination (MMSE), and HDS-R on the same day. The scores for the robot test (robot scores) were calculated by a human. To demonstrate correlations between scores, we used the Spearman's rank-correlation method. Between-group comparisons of scores were assessed using the nonparametric Kruskal-Wallis test, with multiple comparison corrections performed using the Steel-Dwass procedure.

Results

There were no dropouts in the robot test. The robot scores correlated significantly with the scores of the MMSE (rho=0.814, $p<0.001$) and HDS-R (rho=0.855, $p<0.001$). Moreover, robot scores in the AD group were significantly lower than in the MCI ($H=2.93$, $p<0.001$) and HC ($H=1.28$, $p=0.009$) groups.

Received August 30, 2019; accepted November 26, 2019.

Correspondence to: Jumpei Maruta
Department of Neuropsychiatry, Osaka City University Graduate School of Medicine,
1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585, Japan
Tel: +81-6-6645-3821; Fax: +81-6-6636-0439
E-mail: d16mb90s42@st.osaka-cu.ac.jp

Conclusions

Although the interviewer is a general-purpose interactive robot, it is able to talk and get the necessary information to detect poor cognitive function in elderly participants.

Key Words: Interactive robot; Alzheimer's disease; Screening; Neuropsychological examination

Introduction

Early detection of dementia is important because it increases the potential for an effective intervention. For example, interventions such as drug therapy are most effective when they are started early[1]. Brodaty et al reported that neuropsychological examinations such as General Practitioner Assessment of Cognition, Mini-Cog, and Memory Impairment Screen were suitable for dementia screening in general practice[2]. However, since these need a human examiner and have labor-associated costs, it is sometimes difficult to screen dementia due to lack of human resources. Furthermore, the elderly people who undergo cognitive function examinations may feel psychologically burdened by human examiners. Kobayashi et al showed that, due to psychological burden, elderly people may not perform to their full ability in examinations conducted by a human compared to examinations performed using a personal computer[3]. Therefore, automatic and inexpensive tests that do not require a human examiner at the time of testing are needed to screen for dementia.

Previous studies of automated dementia screening have mainly focused on developing screening software for personal computers or tablet devices. However, as pointed out by Kobayashi, the use of a mouse is often difficult for the elderly[3], and examinations using a touch panel is currently preferred. Tsuji et al reported a computer-based test that conducts a neuropsychological examination for elderly subjects with dementia using a touch panel[4]. Onoda and Yamaguchi developed the revision of Cognitive Assessment for Dementia, iPad version (CADi2) and used it for primary mass screening in community-based medical facilities[5].

Recent advances in robot technology also have the potential to enable automated neuropsychological examinations. Several studies have demonstrated the usefulness of robot therapy to treat the behavioral and psychological symptoms of dementia. Gustafsson et al reported that their cat robot increased well-being and quality of life for some individuals with dementia[6]. Treatment with PARO (Daiwa House Industry, Osaka, Japan) a baby seal-like robotic pet was reported to be effective for stress reduction[7] and resulted in reductions in the use of psychoactive medications in elderly people with dementia[8]. In the MARIO Project, it has been reported that the companion robot MARIO may be a useful tool in mitigating depression and loneliness for people with dementia[9]. It is currently unknown whether touch panel devices or conversational robots are more effective for neuropsychological examinations. However, considering robots have been used for elderly patients with dementia as mentioned above, robots likely have the potential to greater reduce the psychological burden than touch panel devices.

We hypothesized that a general-purpose interactive robot would be helpful to assess cognitive function because of the possibility of reducing medical expenses and psychological burden in patients. Hence, we have developed a neuropsychological examination auto-interviewing robot using a general-purpose interactive robot.

In this study, we present our neuropsychological examination auto-interviewing robot and

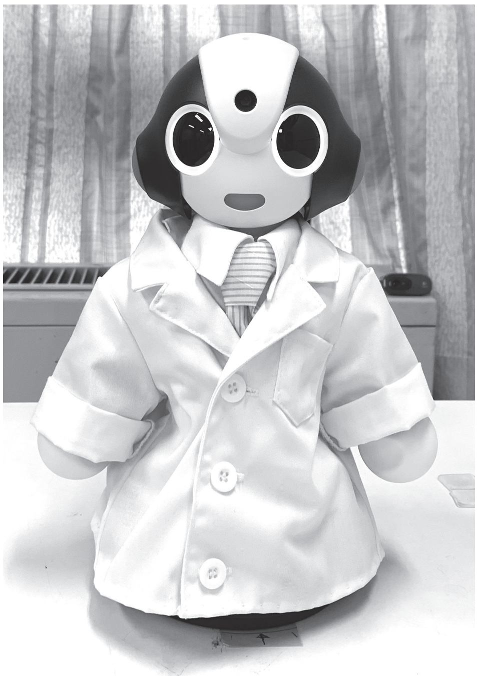

Figure 1. Image of SOTA with white coat.

investigate whether the robot can communicate with and get the necessary information to detect poor cognitive functioning from healthy and cognitive impaired elderly participants.

Methods

Robot architecture

In this study, we used a general-purpose humanoid interactive robot, SOTA (Vstone Co. Ltd., Osaka, Japan), as shown in Figure 1. SOTA has been previously used in a study investigating the use of robots for recreational activities in nursing homes[10]. SOTA measures 280 mm in height and weighs 800g. We used the NTT Data Cloud Robotics Platform to implement SOTA's communication and testing program.

The main function of SOTA was to record the participant's voice and synthesize speech. SOTA communicated with a participant via a loop incorporating the following steps: (a) participant speaks to SOTA; (b) SOTA records the participant's voice with its microphone; (c) voice data is transferred to the NTT Data Cloud Robotics Platform; (d) in the cloud platform, voice data is converted to text data; (e) a response to the text data is generated according to the scenario; (f) an audible response is synthesized; (g) the synthesized audible response is transferred to SOTA; and (h) SOTA responds with a synthesized voice.

Particularly, in step (e), when SOTA received an unexpected response (or no response), SOTA was designed to nod in response and continue to the next question. When the correct reply was received, SOTA repeated the answer.

Robot test

The experiment was conducted in Osaka City Kosaiin Hospital after a baseline examination including a cognitive function test. If for any reason it was not able to be conducted on the same day as the baseline examination, the experiment was conducted as quickly as possible (i.e. within a

Table 1. The robot's brief script and instructions for administration of the test

	Score	Script (originally in Japanese)	Instructions for administration
		I' m SOTA.	Introduction for test.
Q1	1	Tell me your birthday.	
Q2	1	How old are you?	Within 1 year of the patient's correct age
Q3	3	Remember three objects. Cherry, Cat, Train.	Then asks the patient for all three after the robot has said them. Give a point for each correct answer.
Q4	2	Remember three or four numbers, and say them in reverse order.	If the participant answered the three numbers correctly, then proceed to four numbers.
Q5	2	Serial sevens (starting from 90).	83, 76 (One point for each correct)
Q6	6	Recall the 3 objects repeated above.	Give 2 points for each correct. If a patient could not answer completely, give hints and ask one more time. Give 1 point for each correct with a hint.
Q7	5	Remember the following short story: "Yesterday I sent a flower to the teacher who had a reunion in Kyobashi, Osaka."	Give 1 point for each correct word (total five words in Japanese) and subtract 5. In case of a negative score, it is assumed to be 0 points.
Q8	3	Look for similarity between; "an orange and apple", "a bicycle and bus", "a drum and whistle".	Give 1 point for each correct word. "Fruit" "Vehicle" "Musical instrument"
Q9	3	What is the date/month/year?	One point for each correct
Q10	1	Where are we?	"Kosaiin" or "Hospital"
Q11	5	Recall the previous short story.	Give 1 point for each correct word and subtract 5. In case of a negative score, it is assumed to be 0 points.
Q12	1	Say my name.	"SOTA"

month), to avoid any changes to the participant's cognitive abilities between experiments. Participants were guided to the room and introduced to SOTA. We asked the participant to follow any instructions given by SOTA but did not give any advice such as "how to use a robot" during the test.

We created a scenario in which a cognitive function test, with a duration of 10 minutes, was conducted by SOTA (Table 1) in Japanese. The test scenario was developed with reference to the Hasegawa Dementia Scale Revised (HDS-R), which is a validated screening test for dementia[11], and is widely used in Japan. Because of restricted nature of robot conversation, we modified the HDS-R to be suitable for a robot. In particular, recalling 5 objects and generating vegetables were removed. Instead of them, Q7 (immediate recall of short story), Q8 (conceptualization), and Q11 (delayed recall of short story) were added. Furthermore, to get used to the dialogue with the robot, Q1 (birthday) was added as the first question, which is relatively easy to answer even for participants with dementia. Q12 (recall robot's name) was added to verify whether cognitive function can be tested in a natural conversation. To minimize learning effects, we modified questions relating to delayed recall and changed the starting number for the serial sevens so that it differed from the original HDS-R.

After the robot test, a human graded the recorded interview according to a manual (Table 1). The score distribution was determined with reference to HDS-R.

Participants

A total of 26 participants enrolled in the study and completed a baseline examination, including the Mini-Mental State Examination (MMSE) and the HDS-R conducted by a clinical psychologist.

Participants were either outpatients at Osaka City Kosaiin Hospital, Japan, from April 2017 to December 2017, or patients' spouses. Participants were assigned to one of three groups: mild to moderate Alzheimer's disease (AD), mild cognitive impairment (MCI), or healthy controls (HC).

The AD group met the following inclusion criteria: (a) met the National Institute of Neurologic, Communicative Disorders and Stroke, AD and Related Disorders Association, Alzheimer's criteria for probable AD[12]; and (b) scored 11 or more points on the MMSE[13]. MCI patients were defined as having complaints about cognitive impairment, and observable cognitive decline, but whose basic daily living functions were normal. They met the revised MCI criteria[14]. The HC group met the following inclusion criteria: (a) had never been diagnosed with dementia at any hospital; (b) were able to live independently; and (c) had no complaints of memory loss.

We obtained written informed consent from all participants. The study protocol was approved by the Ethics Committee of Osaka City Kosaiin Hospital in accordance with the declaration of Helsinki.

Statistical analysis

Analyses of the demographic and clinical characteristics were performed with Easy R (EZR) (Saitama Medical Center, Jichi Medical University, Saitama, Japan)[15]. EZR is a graphical user interface implemented in R 2.13.0 (R Foundation for Statistical Computing, Vienna, Austria). Between-group comparisons of MMSE, HDS-R and age were assessed by using the nonparametric Kruskal-Wallis test, with multiple comparison corrections performed using the Steel-Dwass procedure. Categorical variables such as sex were compared with Fisher's exact tests. To demonstrate correlations between test scores produced by the robot and humans, we used the Spearman's rank-correlation method. Alpha values less than 0.05 were considered significant.

Results

Acceptance of the robot

No participants dropped out of the robot tests. The participants did not get angry with or confused by our robot enough to prevent the continuation of the robot test.

Demographic characteristics & baseline examinations

For the AD, MCI, and HC groups, sex ratios (female/male) were 4/8, 0/4, and 9/1; the median [first quartile point (Q1), third quartile point (Q3)] ages were 79.00 [74.25, 84.25], 75.50 [73.00, 78.50] and 73.50 [70.25, 75.00] years old; the median MMSE points [Q1, Q3] were 21.50 [20.75, 22.00], 25.00 [25.00, 25.25] and 26.50 [24.50, 27.75]; the median HDS-R points [Q1, Q3] were 18.00 [14.00, 19.00], 23.00 [19.50, 26.25] and 29.00 [28.25, 30.00].

There was a significant group difference of sex ratio among the AD, MCI, HC groups ($p=0.002$, Fisher's exact tests). There were fewer male participants in the HC group compared to AD and MCI group.

The AD group were significantly older than HC group ($H=2.40$, $p=0.044$). The median age of the MCI group was not significantly different compared to the AD ($H=0.86$, $p=0.668$) and HC ($H=1.28$, $p=0.403$) group.

The AD group had significantly lower MMSE scores than both the MCI group ($H=2.78$, $p=0.015$) and HC group ($H=3.21$, $p=0.004$). There was no significant MMSE score difference between the MCI and HC groups ($H=1.00$, $p=0.576$).

The AD group had significantly lower HDS-R scores than the HC group ($H=3.98$, $p<0.001$). MCI group tended to have lower HDS-R scores than HC group ($H=2.30$, $p=0.056$).

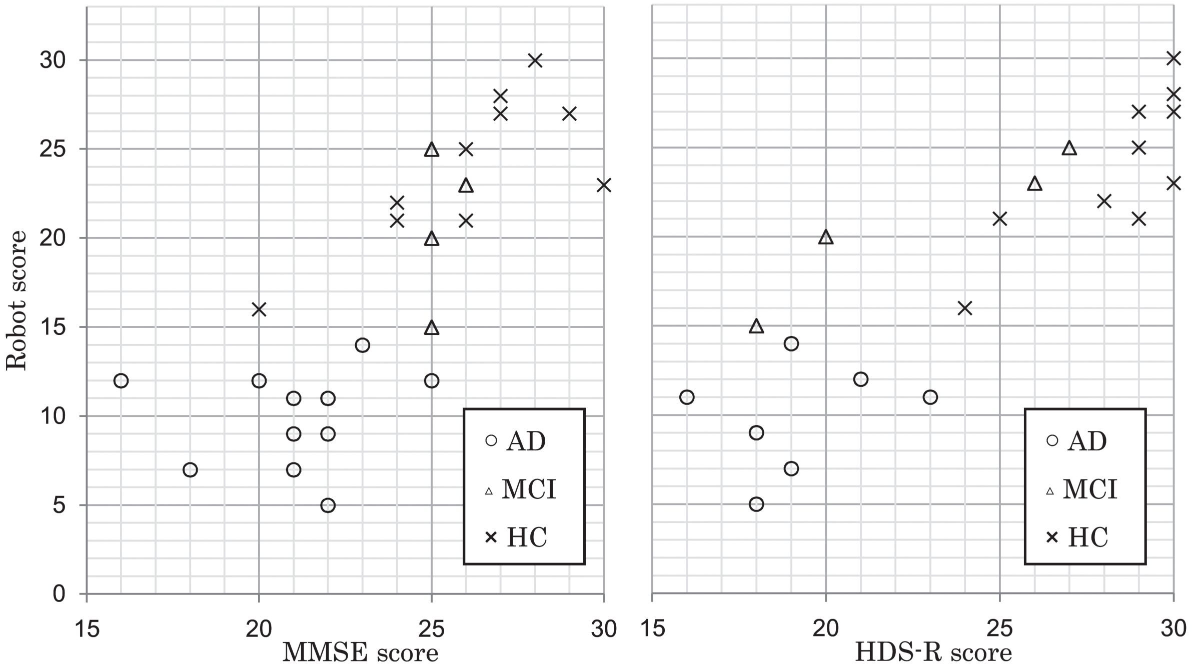

Figure 2. Score correlation between robot scores and neuropsychological examinations. Abbreviations: AD, Alzheimer's disease; MCI, mild cognitive impairment; HC, healthy control; MMSE, Mini-Mental Scale Examination; and HDS-R, Hasegawa Dementia Scale Revised.

Table 2. Average scores of each question of robot test

Question (max point)	Average score±SD			p value
	AD	MCI	HC	
Q1. Birthday (1)	1.00±0.00	1.00±0.00	1.00±0.00	NaN
Q2. Age (1)	0.83±0.39	1.00±0.00	1.00±0.00	0.308
Q3. Repeating 3 words (3)	2.33±0.78	2.75±0.50	2.40±0.84	0.649
Q4. Digits backward (2)	0.92±0.79	1.25±0.96	1.30±0.67	0.488
Q5. Calculation (2)	1.42±0.67	2.00±0.00	1.50±0.71	0.298
Q6. Recalling 3 words (6)	0.25±0.62	3.00±2.58	5.30±1.06	<0.001
Q7. Immediate recall of short story (5)	0.08±0.29	3.00±1.41	2.80±1.40	<0.001
Q8. Conceptualization (3)	2.00±1.04	2.25±0.96	2.90±0.32	0.055
Q9. Orientation to time (3)	0.67±0.98	2.00±0.00	3.00±0.00	<0.001
Q10. Orientation to place (1)	0.50±0.52	0.75±0.50	1.00±0.00	0.028
Q11. Delayed recall of short story (5)	0.00±0.00	1.75±1.71	1.70±1.95	0.015
Q12. Recall robot's name (1)	0.00±0.00	0.00±0.00	0.10±0.32	0.467
Total (33)	10.00±2.63	20.75±4.35	24.00±4.19	<0.001

P values represent group differences among AD, MCI, and HC groups with the Kruskal-Wallis test. Abbreviations: AD, Alzheimer's disease; HC, healthy control; HDS-R; Hasegawa Dementia Scale Revised; MCI, mild cognitive impairment; MMSE, Mini-Mental Scale Examination; NaN, not a number; and SD, standard deviation.

Robot scores

Figure 2 shows score correlations between the robot test and the neuropsychological examinations MMSE and HDS-R. There was a significant correlation between the robot scores with the MMSE (rho

=0.814, $p<0.001$) and the HDS-R scores (rho=0.855, $p<0.001$).

Robot scores in the AD group (median, 11.0) were significantly lower than those in the MCI group (median, 21.5; H=2.93, $p<0.001$) and HC group (median, 24.0; H=3.97, p=0.009). There was no significant difference in robot scores between the MCI and HC groups (H=1.28, p=0.407).

Table 2 shows the average scores of each question of the robot test. There were significant group differences in Q6 (recalling 3 words), Q7 (immediate recall of short story), Q9 (orientation to time), Q10 (orientation to place) and Q11 (delayed recall of short story). All participants in the 3 groups answered Q1 (birthday) correctly. Only HC participants answered Q12 correctly.

Discussion

We performed a pilot study for the use of a general-purpose interactive robot to screen for dementia. The robot conducted a modified HDS-R, which was completed by all participants. There were significant correlations between the robot scores and MMSE and HDS-R scores conducted by a human psychologist. Previous studies have supported the usefulness of the MMSE and HDS-R as screening tests for Alzheimer's disease[11,16]. Moreover, robot scores of 14 or less were obtained only in patients with dementia and not in those with MCI or HC. These results suggest our robot scores are useful to detect poor cognitive functioning in healthy and cognitive impaired elderly participants. Therefore, our novel robot was able to talk to and get the necessary information from elderly patients to make such a detection.

It is worth noting that there were no dropouts in the robot test despite some of the participants included having mild to moderate AD. This supports the idea that patients with mild to moderate AD can accept a general-purpose interactive humanoid robot in well-structured interactions as an interviewer. Furthermore, although the robot represents an inspector, it appears that there was no psychological burden that would make psychological examination difficult to complete in this study. However, because our study focused on screening and was not targeted at patients with severe dementia, it is still unclear whether a patient with severe dementia could communicate with our robot. A previous study has shown that a humanoid robot can be used with institutionalized patients with advanced dementia, but in that study, unlike here, the robot was remotely operated[17].

In the robot test, the breakdown of the subject's score was comparable to that of the conventional test. The loss of recent memory (Q6, recalling 3 words; Q11, delayed recall of short story) and orientation (Q9, orientation to time; Q10, orientation to place) was striking while short-term memory loss (Q3 repeating 3 words; Q4, digits backward) was relatively preserved in AD and MCI patients. This preservation of short-term memory is in line with previous studies showing that short-time memory is intact in patients with a damaged medial temporal lobe, such as AD patients[18,19].

There are three important tasks revealed by this study regarding our modified HDS-R, and it will be necessary to improve the content of the questions for future research. First, the administration time may be long. In this study, we tested various types of questions to investigate what kind of question is best suited for a robot. Therefore, the robot test can be shortened in the future. Second, internal consistency could be improved. Some questions (Q1-5, Q12) were either too easy or too difficult to distinguish AD or MCI participants from HC. Third, participants in the MCI and HC groups could not be clearly differentiated by our robot test. The cause of this could be the small sample size of the experiment, or the difficulty of the robot test. The difficulty of the robot test was designed using HDS-R as a reference, which is designed to screen for dementia, not MCI. Therefore, the robot test was considered to be too easy to distinguish MCI from HC.

Compared to the Montreal Cognitive Assessment (MoCA), which has been reported as a useful screening tool for MCI[20], our test lacks a trail making test because our robot can only interact verbally in this study. When MMSE and a trail making test were combined, the accuracy and precision of mild dementia screening was reported to improve[21]. Similarly, to accurately screen for MCI we might need to include a trail making test in future studies. Some commercially available robots such as PEPPER (Soft Bank Corp., Tokyo, Japan) have touch panels to aid in communication, and therefore could easily incorporate a trail making test into our test.

There are three advantages to our robot test compared to the paper based conventional test. First, no expert is required to conduct the cognitive function tests. Second, we can standardize the method of conducting cognitive function tests. Third, since the recorded interviews are scored, labor-associated costs are minimized. Bearing in mind privacy issues, recorded interviews can be sent electronically and scored anonymously at a centralized location. Considering that SOTA has been studied for use as a recreational robot in nursing homes, we believe that our robot is better suited to support dementia screening in nursing homes that lack the necessary human resources.

This study has several limitations that should be considered when interpreting the data. First, the number of participants was small. There was no precedent for the application of an interactive robot to a neuropsychological examination. Therefore, we considered that the number of cases was appropriate as an initial pilot study aimed at optimizing the protocol for future studies. Second, the sex ratio and age of participants was biased, and it is possible that sex and age may influence scores in the robot test. However, strong correlations between robot scores and the MMSE (or HDS-R) suggests that the robot scores are not affected by the sex and age of the examinee. Third, responses to the robot were not quantified, therefore we could not demonstrate a statistical association between the severity of dementia and the responses to the robot. Finally, there is a possibility that the learning effect occurred, because MMSE and HDS-R were performed before the robot test. It is possible that the task was performed appropriately as a result of having experienced the MMSE and HDS-R, even if the instruction by the robot was inadequate. However, most of the instructions in the robot test were modified for the robot conversation from original HDS-R. Furthermore, to minimize learning effects, we modified questions relating to delayed recall and changed the starting number for the serial sevens. Therefore, we assume the learning effect did not significantly affect the robot score or the interactions between participants and robot in this study.

Conclusions

Even though the interviewer is a general-purpose interactive robot, our novel robot was able to talk to and get the necessary information from elderly participants to detect poor cognitive functioning. Our pilot study suggests that a general-purpose interactive robot could be a promising support tool for dementia screening. Although our robot currently requires human scoring, it can still reduce human cost because a non-specialized human can effectively score recorded interviews from robot tests in their spare time. We expect that in the future, interactive robot tests will become a valid and reliable tool for clinicians when screening for dementia in nursing homes that lack appropriate human resources.

Acknowledgements

All authors have no COI to declare regarding the present study.

We would like to thank Editage (www.editage.jp) for English language editing. We are also grateful to Mai Horimoto, Yukiho Watai, and Masayo Kobayashi for conducting neuropsychiatry examinations.

References

1. Atri A. Effective pharmacological management of Alzheimer's disease. Am J Manag Care 2011;17:S346-355.
2. Brodaty H, Low LF, Gibson L, et al. What is the best dementia screening instrument for general practitioners to use? Am J Geriatr Psychiatry 2006;14:391-400.
3. Kobayashi S. Neuropsychological test using personal computer. Japanese Journal of Neuropsychology 2002:18: 188-193. (In Japanese)
4. Tsuji M, Higashi Y, Iwase Y, et al. Neuropsychological examination of elderly subjects with dementia and mentally healthy elderly using a touch panel of a personal computer. Journal of Life Support Engineering 2009; 21:158-163. (In Japanese)
5. Onoda K, Yamaguchi S. Revision of the cognitive assessment for dementia, iPad version (CADi2). PLoS One 2014;9:e109931.
6. Gustafsson C, Svanberg C, Müllersdorf M. Using a robotic cat in dementia care: a pilot study. J Gerontol Nurs 2015;41:46-56.
7. McGlynn S, Snook B, Kemple S, et al. Therapeutic robots for older adults: investigating the potential of Paro. HRI'14 Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction. ACM, NY, USA, 2014. pp. 246-247.
8. Petersen S, Houston S, Qin H, et al. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55: 569-574.
9. D'Onofrio G, Sancarlo D, Raciti M, et al. MARIO project: validation and evidence of service robots for older people with dementia. J Alzheimers Dis 2019;68:1587-1601.
10. Kuwahara N. Assessing the use of communication robots for recreational activities at nursing homes. European Union Digital Library 2016;16(7):e4.
11. Imai Y, Hasegawa K. The Revised Hasegawa's Dementia Scale (HDS-R) - Evaluation of its usefulness as a screening test for dementia. Hong Kong Journal of Psychiatry 1994;4:20-24.
12. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939-944.
13. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189-198.
14. Artero S, Petersen R, Touchon J, et al. Revised criteria for mild cognitive impairment: validation within a longitudinal population study. Dement Geriatr Cogn Disord 2006;22:465-470.
15. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013;48:452-458.
16. Kim KW, Lee DY, Jhoo JH, et al. Diagnostic accuracy of mini-mental status examination and revised Hasegawa Dementia Scale for Alzheimer's disease. Dement Geriatr Cogn Disord 2005;19:324-330.
17. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, et al. Social robots in advanced dementia. Front Aging Neurosci 2015;7:133.
18. Baddeley AD, Warrington EK. Amnesia and the distinction between long- and short-term memory. Journal of Verbal Learning and Verbal Behavior 1970;9:176-189.
19. Milner B. Disorders of learning and memory after temporal lobe lesions in man. Clin Neurosurg 1972;19:421-446.
20. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53:695-699.
21. Wouters H, Appels B, van der Flier WM, et al. Improving the accuracy and precision of cognitive testing in mild dementia. J Int Neuropsychol Soc 2012;18:314-322.