Liu et al., Gerontechnology (2022) 21:1
https://doi.org/10.4017/gt.2023.21.1.794.02

Humanoid robot teleoperation reminiscence group therapy for older adults with dementia: A controlled trial study

XiangYu Liu MA[a,*], Ryuji Yamazaki PhD[b], Jie Yang MA[a], Hiroko Kase PhD[c]

*[a]Graduate School of Human Sciences, Waseda University, Japan; [b]Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Japan; [c]School of Social Science, Western Sydney University. Australia; *Corresponding author: dukeedison@fuji.waseda.jp*

Abstract

Background: According to the Ministry of Health, Labor, and Welfare of Japan, there will be a shortage of 270,000 nursing staff in Japan by 2025. The Japanese government has also proposed using intelligent technology to alleviate this problem. Introducing robots into nursing facilities is an essential solution to the labor shortage.
Objective: This study attempts to find the differences between humanoid robots in replacing caregivers through linguistic and emotional comparisons with a controlled trial. The actual use of the humanoid robot Pepper for dementia in older adult care facilities was conducted. In addition to the common scales and linguistic analysis, an interdisciplinary approach was used to increase the dimension of facial expression recognition to better represent the differences in communication between elderly people with dementia and different objects(human or robot).
Method: We conducted a controlled trial for 8 participants into two groups, G1 (average age=85.75, SD±3.7) and G2 (average age=89, SD±2.73). G1 was divided into face-to-face Reminiscence Group Therapy (RGT) and G2 into robot-teleoperated RGT for six sessions. The interventional experiment lasted from December 4, 2020, to January 12, 2021. The differences in Mini-Mental State Examination (MMSE), Neuropsychiatric Inventory-Nursing Home Version (NPI-NH), and Dementia Behavior Disturbance (DBD) scales between the two groups before and after the intervention experiment was adopted. Linguistic analysis (including sentence-ending particles and Entropy) using independent sample t-test. Results of the trend using a facial emotional recognition tool Emo-Rec App was introduced to compare differences.
Results: The two intervention methods have no obvious differences in MMSE ($t(8)=0.156$, $p=0.881$), NPI-NH($t(8)=0.380$, $p=0.717$), and DBD($t(8)=-1.001$, $p=0.355$) scales. Utterance($t(8)=2.434, p=0.022$) and sentence-ending particle “ne” ($t(8)=2.313, p=0.029$) existed statistical difference. Mood changes of the humanoid robot group have improved intuitively (Emo-Rec point 10.387 to 8.531 on average).
Conclusion: The linguistic results for the teleoperated humanoid robot group were somewhat inferior. On the opposite, the mood of the robot group was improved by a better trend.

Keywords: humanoid robot, teleoperation, dementia, RGT, nursing home, nursing technology

Introduction

According to the population estimation data recently released by the Ministry of Health, Labor and Welfare of Japan, the number of older adult people over 65 years old in Japan has increased by 320,000 compared with the previous year, reaching 35.88 million, accounting for 28.4% of the total population (Ministry of Health, Labor and Welfare, 2020). An increasing proportion has been expected in the future, reaching 30.0% in 2025 and 35.3% in 2040 (MHLW, 2020).

The aggravation of the aging problem has multiple adverse effects on the Japanese economy and society. By sustainable increasing, the burden of pensions and medical care leads to tremendous pressure the Japanese finances face. All those older adults born during the first baby boom after World War II will be over 75 years old in 2025, and there will be a shortage of approximately 270,000 nursing staff (MHLW, 2020). In response to the aging population, the Japanese government has attached importance to developing Intelligent Technology (IT) for older adult care and has promoted it as an economic industry. The pace of IT development has also become one of the Japanese advantages. Japan uses intelligent technology to realize comprehensive care for older adults. The medical and nursing robot market is expected to be 534 billion yen in 2025, increasing 2.21 times compared with 2020 (Fuji Keizai, 2021). In this regard, the future of the ro-

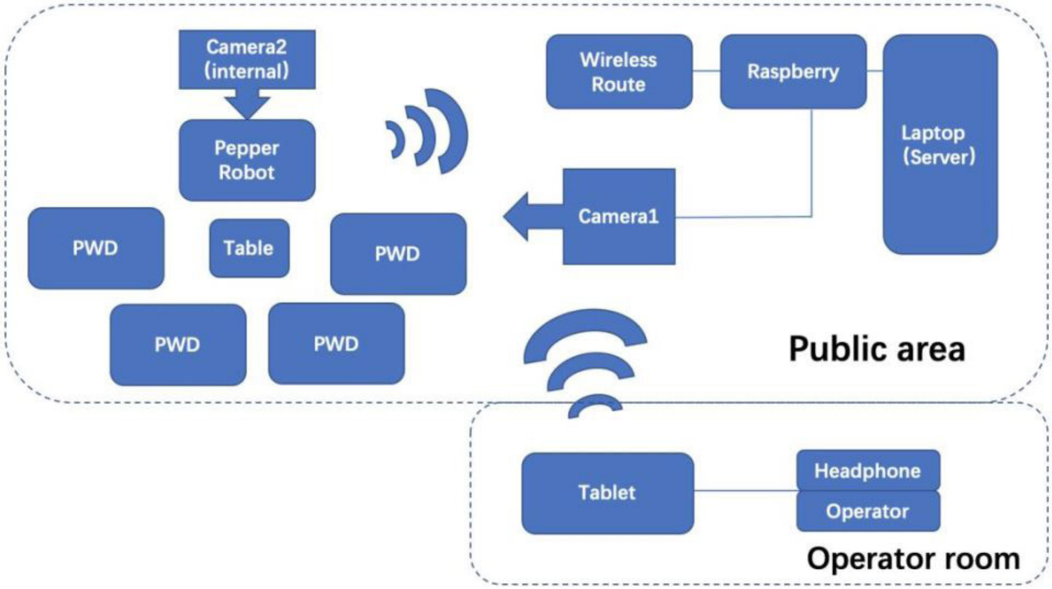

Figure 1. Design of experiment

bot economy will be promisingly growing.

In the robotics industry, the earliest Paro seal robot design started in 1993 and came out in 2001. It pioneered research on the care of older adults with dementia and the relief of symptoms such Behavioral and Psychological Symptoms of Dementia (BPSD) (Moyle, Jones, Cooke, Dwyer, Sung & Drummond, 2017). Since then, the concept of a Socially Assistive Robot has formally appeared. Robots used in older adult care and older adult care with dementia have gradually been developed and are relatively widely used (Abdi et al., 2018) in facilities worldwide. As robot industries developed, such as semiconductors and manufacturing, Socially Assistive Humanoid Robot (SAHR) has gradually become used daily. Compared with previous pet-type robots, SAHR can achieve more symptomatic applications for specific manifestations such as BPSD. For example, relieving the peripheral symptoms of BPSD through Human-Machine Interaction helps Person With Dementia (PWD) return to society through counseling and communication (Liu, 2020).

Pepper is the world's first social humanoid robot to recognize faces and basic human emotions. Pepper was optimized for human interaction and can engage with people through conversation and his touch screen (Softbank, 2019). Pepper is highly valued for future expansion and research as a comparatively mature household and enterprise-integrated robot. It has provided convenience for social operations regarding distance education and voice calling in recent years. Engineers continue to develop excellent applications via Pepper Innovation Challenge to play a role in more work scenarios. During the COVID-19 pandemic, we are increasingly aware of the importance of virtual interaction. Virtual interaction between humans and robots can be applied to more scenarios than just the exclusive domain of young people. Dialogue communication through electronic devices can get closer and closer to face-to-face communication, and the boundaries between virtual and reality may reduce. Virtual reality is not limited to both parties in touch using electronic devices to access the virtual space but also includes one party using input-type electronic devices to control robots and communicate remotely. For the broader application of new technologies, what are the remaining shortcomings of teleoperated robots in replacing the labor force of PWD care are the questions that need to be clarified in this study. This study used Pepper as a SAHR to implement teleoperated functionality and applied it to PWD with a controlled trial study.

Background

The research surrounding Pepper has many similar applications; for example, using Cognitive Stimulation Therapy (CST) to help PWD's rehabilitation training and help their cognitive rehabilitation (Carolis et al., 2020). In addition, there are some other short-term contacts and good feedback on Pepper's intervention research cases. However, Pepper's previous AI dialogue system evaluations in senior facilities were unsatisfactory (Kase, Yamazaki, Zhu & Nishio, 2019). The early developed version was too austere for Human-Machine Interaction, and the dialogue content was too straightforward for long-term communi-

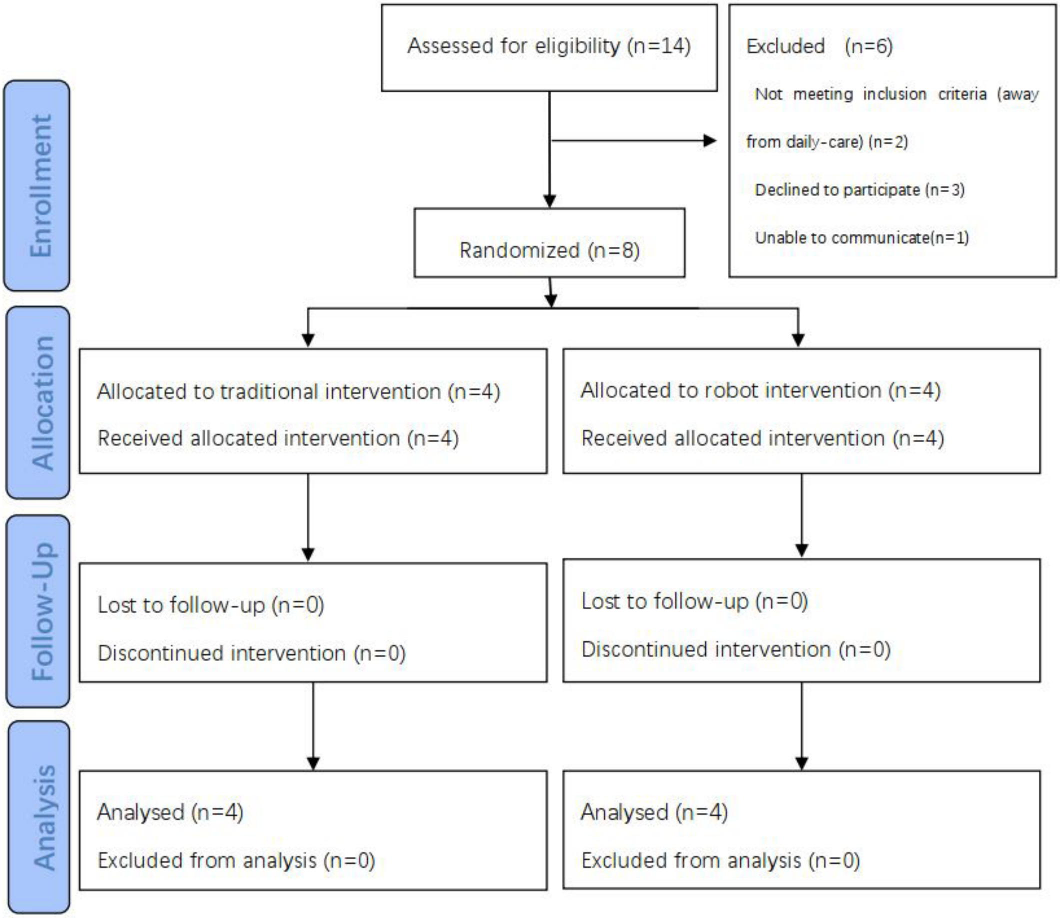

Figure 2. Flow chart of the study

cation. However, Pepper's development environment can support the function of remote video calls from the operator so that it can function similarly to Giraff, Telenoid, and any other Socially Assistive Robots through software and hardware redevelopment (Moyle et al., 2017).

Through Social Assistance Robot (SAR) scoping reviews, we learned that research reports on older adult care and dementia-corresponding robots focus on pet-type robots(Abdi et al., 2018). Still, reports on Social Assistant Humanoid Robot (SAHR) are relatively fewer. A report pointed out that judging how humanoid robots can help dementia BPSD more effectively could be evaluated in terms of human imitation, social interaction, and companionship (Liu, 2020), the new model SAHR will bring better interventional results. Research on Giraff asked participants' families to communicate with PWD (n=5) through robots and supporting software to prove whether the Social Assistant Robot can be used in long-term nursing, which led to a satisfying result. However, it also pointed out that there were limitations to using this type of robot in older adult care facilities. The software had a particular learning cost for family members, and there were specific software and hardware compatibility and network stability issues during the experiment. In addition, Wendy Moyle also pointed out that the evaluation of facial expressions in the study of PWD would be an excellent non-verbal expression evaluation method (Moyle et al., 2014).

Instead of using AI dialogue, the Teleoperated format seems to be the best option at the moment. Besides, we will compare and discuss another dimension, such as Emo-Rec, for facial expression recognition. There was also a study to verify whether deep learning facial emotion recognition applications can correctly recognize the emotion of PWD through the judgment of older adult facility staff and nurses, and came to a positive conclusion (Liu, 2021).

In summary, we need to redevelop Pepper to ensure that Pepper inherits the benefits of SAR in the intervention experiments with PWD. From

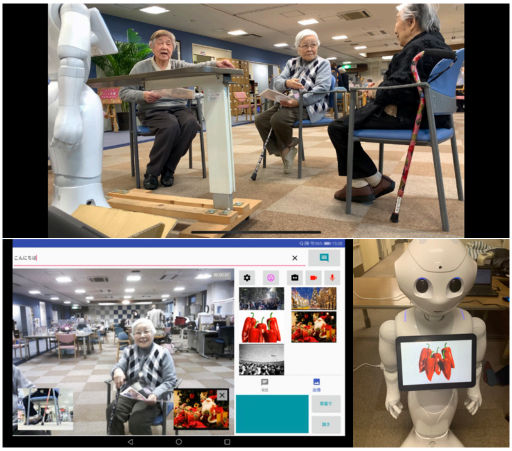

Figure 3. Vision in Camera 1, Camera 2 (tablet). Picture on pepper

the perspective of labor-saving, the RGT method is currently more feasible, reducing the pressure on caregivers and providing more labor opportunities for the initial older adult population. When elder adults who can work become operators, they need to read the simple manual and then easily communicate with PWD according to the theme of the RGT.

Evaluating the effects of robotic intervention experiments through linguistics is an advanced approach. This method can objectively demonstrate the impact of intervention experiments through quantitative data analysis. A linguistics analysis in the RGT intervention experiment using a Telenoid robot to PWD (n=6) pointed out that PWD has a higher average number of human conversations than the robot. Nevertheless, there was no significant difference in language complexity and content richness. In addition, one of the difficulties in evaluating teleoperated robots is that quantitative analysis should be introduced. According to literature reviews and previous studies, the number of participants in this type of research ranges from two to three to a dozen and is still at a relatively early stage (Kase et al., 2019).

Based on the above research and literature review, the Pepper robot developed by Softbank was used in this intervention experiment. A certain degree of development was conducted to enable it to conduct remote video communication. The findings are based on observation-based scales, adding linguistics and emotion recognition as objective indicators.

The study

Design

In elderly care facilities, the RGT method has been used widely to provide better dementia BPSD relief services for the elderly in daily care. In Japan, the work pressure on nursing staff is high due to the labor shortage. For the broader application of new technologies, what are the remaining shortcomings of teleoperated robots in replacing the labor force of PWD care are the questions that need to be clarified in this study. This study used Pepper as a SAHR to implement teleoperated functionality and applied it to PWD with a controlled trial study.

We conducted an RGT intervention study using Pepper in an older adult care facility in Nakameguro District, Tokyo, Japan. The experiment used more wireless connections instead of wired connections to ensure the cleanliness of the facility and avoid the risk of causing older adults to fall accidentally. Due to the RGT therapy viewing angle of the robotic camera, 3-4 participants per RGT would be an ideal configuration. Eight participants from the facility were screened and selected by limited conditions and divided into two groups conventional RGT and robot RGT (*Figure 1*). limited conditions included no willingness to withdraw from the daily care service during the trial, essential communication and cognitive ability, and willingness to participate in the trial.

We completed the screening and grouping of subjects with the help of the facility's daily care supervisor. 8 out of 14 participants were randomly selected to participate in the study, and the participants' criteria were daily care users in the facility and had varying degrees of dementia. The supervisor randomly selected eight sticky notes with participant numbers and placed them in the robot and traditional communication groups (*Figure 1*). A t-test was used on the questionnaire scale, and an independent samples t-test was used in the linguistic analysis to compare differences in the intervention experiments. Finally, the results of the AI tool are summarized, and the results are discussed using trend plots.

Interventions

The setting of RGT is one-to-many. With the help of staff in the older adult care facility, we set up a group of 4 PWDs, an interventional experiment of two groups (n=8). The average MMSE score was 18.375 (SD±7.60), and participants were considered to have mild to moderate cognitive impairment (MCI). Among the 8 participants, one could not generally speak after tracheotomy and could not evaluate the effect through linguistic analysis. We will add the dimension of facial emotion recognition to assess the specific differences between the two intervention methods.

We used the Rasberry Pi as a remote control transfer station, which can help the operator's tablet connect to the speaker and the second camera. Then we attached the Rasberry Pi to the server and the wireless router. The server will provide an

Figure 4. Nursing recreation memory card for RGT

IP address to communicate with the operator side so that the pepper side can communicate. These three parts of the devices are connected through wireless networks (*Figure 2*). Through the tablet, we may display any required pictures on the display screen of the Pepper (*Figure 3*).

The experiment was conducted in a public residential facility for seniors. Following the facility's schedule, we divided the participants into two groups. One communicated with a human coordinator, and another communicated with the same human coordinator through Pepper. The operator was a 70-year-old local Japanese volunteer and had more common topics with the participants than young people.

From December 4, 2020, to January 12, 2021, all the participants conducted 6 RGT sessions for 20 minutes. Every session started with greetings and self-introduction to guide participants to remember each other's names. Afterward, the participants needed to watch several nursing recreation memory cards with specific numbers, interact with the cards' content, and refresh the cards for the next session (*Figure 4*). A nursing recreation memory card for RGT is a solution to the problems of field staff who do not know how to communicate with older adults. Two groups of participants would be distributed the same numbered cards at the same sequence of sessions to control variables.

Participants

With the help of the facility staff, we learned the basic information about the participants and divided them into two groups of participants. The first group communicated with a human coordinator (G1, n=4, average age=85.75, SD±3.7). The other group transmitted with the same human coordinator through the humanoid robot Pepper (G2, n=4, average age=89, SD±2.73) (*Table 1*).

Data collection

We completed the MMSE (Mini-Mental State Examination) scale directly with the participants one-on-one (*Table 1*), the MMSE is a widely used questionnaire that reflects the cognitive status of PWD.

With the help of nursing staff in the nursing home, we completed the NPI-NH (Neuropsychiatric Inventory-Nursing Home Version) and DBD (Dementia Behavior Disturbance) questionnaires before the interventional experiment on December 4, 2020 (T0), and one more time on January 12, 2021 (T1) after the interventional experiment. NPI-NH and DBD scales could be able to present assessments for PWD from caregivers including physical condition, cognitive status, and burden of dementia. MMSE, NPI-NH, and DBD scales in English have been added to the appendix.

Then we recorded all the RGT experiments with a voice recorder and a video camera, transcribed them into text, and used them in the subsequent linguistic analysis. By recognizing facial expressions, we divided the emotional changes and trends of every RGT participant. We randomly selected nine video frames for each stage and let the Emo-Rec Application score nine points in the video (*Figure 5*). The random process used SPSS 12 computation to determine the number of seconds in the video and extract the frame.

Ethical considerations

This study has complied with the requirements of the Waseda Ethical Review Committee. (Ethical review number 2019-328) The experimenters explained the experiment content, possible risks, and personal information protection to each participant. All participants signed the consent form for study participation.

Data analysis

For MMSE, NPI-NH, and DBD results, we used independent samples T-test to compare the subjects' cognitive level and nursing status improvement before and after the interventional experiment. We recorded all the conversations

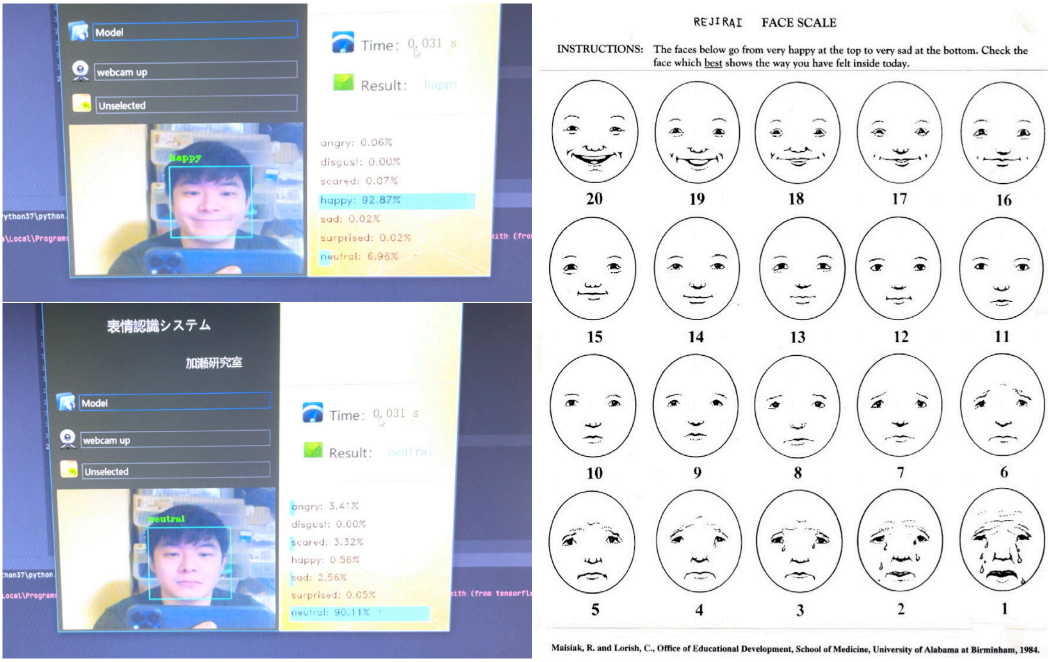

Figure 5. The Emo-Rec application and the face scale

during the session and divided them into two methods: linguistics analysis and emotion recognition analysis using IBM SPSS Statistics Software (R26.0.0.0 64bit version).

Linguistics analysis

1. Utterance and sentence-ending particles

Linguistics analysis includes statistical analysis of sentence-ending particles ("yo", "ne"). For this purpose, we used an open-source text segmentation library, "Mecab" (version 0.996, dictionary ipadic). The frequencies of the sentence-ending particles "yo" and "ne" indicate the softer expressions in daily Japanese conversations as well as natural and relaxed conversations (Den Y,2008). We tried to identify the possibility of using robots as substitutes for people in therapy by independent samples T-test analyzing the differences in tone and wording between PWD in the face of people or robots.

2. Entropy

The other part of linguistic analysis is the analysis of Entropy. The term "Entropy" was first proposed in 1948 and referred to the amount of communication output information whose average value can be measured (Shannon, 1948) (Lash, Rogers & Zoller, et al., 2013). Measuring the Entropy of spoken words can reflect whether the dialogue is compelling and rich in content. N-gram text which is used for measuring the "Entropy" was widely used for evaluating speech and natural language processing tasks (Lodhi et al., 2002) (Suzuki et al., 2017). We calculated the Entropy of 2-gram, 3-gram, and 4-gram extracted from conversation data (*Table 2*).

The formula of Entropy is as follows:

Entropy(xi) is the probability of occurrence for item i. log p(xi) is the amount of information for item i and the overall formula is the average amount for all items.

$$H(x) = -\sum_{i=1}^{n} p(x_i)\log p(x_i)$$

The situation of the participants could be reflected by comparing the number of spoken words, the sentence-ending particles, and the Entropy. Therefore, it is preliminarily to determine whether the participant has achieved the required RGT effect in the intervention experiment (*Table 3*).

Emotion recognition analysis

In this experiment, an Emotion Recognition Application uses a camera to recognize faces and query a database for matching. The Emo-Rec Application judges patients' emotions with dementia through objective data and higher efficiency. In addition, existing studies have compared this type of data with the judgments of elderly facility staff and nurses, confirming part of the reference value of this tool in practical use (Liu, 2021). Face Scale was widely used for patients with emotional difficulties in the past (Lorish & Maisiak, 1987). They were asked to choose the emotions from 1-20 expressions and then give feedback to researchers at that time (*Figure 5*). This study continued this scale of 1-20

Table 1. Comparison of the questionnaire results between traditional RGT and Robot RGT groups (n=8)

		Time	G1(n=4) M±SD	G2(n=4) M±SD	t	p
Age（years）		-	85.75±3.7	89.00±2.37	-1.223	.267
Nursing status		-	2.75±0.957	2.75±0.957	-	-
MMSE [a]		T0	15.25±4.85	17.25±12.21	-.368	.725
		T1	17.50±3.59	19.25±12.23	-.314	.771
		MD	2.00±1.63	1.75±2.75	0.156	.881
NPI-NH [a]	**Severity**	T0	9.25±5.37	4.00±3.55	1.628	.155
		T1	16.75±17.11	8.75±7.88	0.849	.428
		MD	7.50±12.26	4.75±7.67	0.380	.717
	Caregiver Distress	T0	15.75±13.5	5.00±5.30	1.480	.489
		T1	26.75±18.06	17.00±11.34	0.914	.396
		MD	11.00±7.34	12.00±8.75	-0.175	.867
DBD [a]		T0	34.50±29.21	13.75±14.24	1.277	.249
		T1	32.75±29.51	24.25±20.27	0.475	.652
		MD	-1.75±6.55	10.50±23.57	-1.001	.355

Notes: G1 Traditional RGT group, G2 Robot GRT group,T0 before intervention, T1 after intervention, [a] Independent Samples T-test, MD mean difference T1-T0.

to facilitate communication with staff and data comparison. Through the weighted formula, we can convert the collected facial expression data into a scale of 1-20 for analysis.

The formula of Emo-Rec Point is as follows:
Emo-Rec point=(anger*10) + (disgust*10) + (scare*10) + (happy*1) + (sad*20) + (surprise*10) +(neutral*10)

For example, a frame of video with a result of 30% happy and 70% sad would result in an emotional score of 14.3. It can be assumed that the more satisfied the participant's expression, the closer the number is to 1, whereas the sadder it is, the closer it is to 20, and the more neutral it is to 10. The Emo-Rec points allowed us to quantify the emotions of the RGT participants and conduct the subsequent analysis.

We marked nine-time frames in each RGT video by referring to the random sequence given by SPSS 12 Compatible and classified the frames of each video using the Emo-Rec app to derive the proportion of each emotion expressed by participants (Figure 5). After collecting the average Emo-Rec points of each session, we analyzed the difference between the two intervention methods by trend.

Validity and reliability
In this study, we discuss the differences between the two intervention methods. It is not accurate enough to judge only by the scale of participants and facility staff. So we have added two contrasts in linguistics and emotion recognition, trying to discuss the difference between robots and humans in RGT from multiple angles. Counting the speeches (Utterances) in the two intervention experiments can reflect the state of the old adults when they face robots or humans. Linguistic analysis can objectively reflect whether there is a significant difference in the effect when there is a substantial difference in the number of utterances.

In the previous emotion evaluation standards of the face scale, participants were required to express their mood state subjectively. Obtaining data on objectively evaluated expressions seems to be becoming challenging. Under this consideration, relatively objective results would be accepted if the improvement of emotions is measured using deep learning AI. Part of the code of this tool refers to the code on GitHub: https://github.com/oarriaga/face_classification. The detection output results are divided into seven expressions: neutral, angry, contempt, disgust, fearful, happy, and sad. This article uses The data set containing 28709 training samples, 3859 validation data sets, and 3859 test samples, a total of 35887 images containing seven expression categories. The image resolution is 48. ×48. Most images in this dataset have rotations on planes and non-planes, and many images are blocked by occluders such as hands, hair, and scarves. We use the training model integrated into the FER2013 data set (another relatively complete classification database). The classification accuracy rate is up to 74%. The intraclass correlation efficiency (ICC) between the Emo-Rec app and the nursing staff scores is 0.835, which can be considered to have a certain consistency.

Result
In the results of each scale, there was no statistical difference before and after receiving the intervention methods. The results indicated no difference between the two groups regarding the level of cognition and the burden on the caregivers (*Table 1*).

Questionnaire results
Before and after the experiment, we conducted MMSE, NPI, and DBD assessments on the two groups of participants, G1 (average age=85.75, SD±3.7) and G2 (average age=89, SD±2.73). The deterioration of the physical condition of

Table 2. The entropy of 2 groups: G1 traditional RGT group, G2 Robot GRT group, grams the frequency of a phrase composed of consecutive Japanese pronunciations

Session	2-grams		3-grams		4-grams	
	G1	G2	G1	G2	G1	G2
1	3.18453530	3.21223410	3.25836452	3.30589882	3.25384742	3.28126706
2	3.17723179	3.27661087	3.25791083	3.29989639	3.17149157	3.26393343
3	3.22693980	3.15586648	3.27547739	3.22881258	3.27123446	3.30715623
4	3.19637875	3.21461436	3.22427978	3.29568517	3.29933624	3.27608696
5	3.22744593	3.22759142	3.28830967	3.30557775	3.28757958	3.28649584
6	3.24397651	3.24308944	3.25098311	3.24589548	3.18518576	3.27878398

several participants during the experiment led to an increase in the indicators of NPI and DBD. The rest of the participants had different degrees of cognitive improvement (Kitamura, Imamura, Akemi & Iwabashi, 2010). We took the mean difference (T1 minus T0) between the two groups for independent sample T-test analysis to compare the questionnaire results of the participants under different intervention methods (*Table 1*).

From the perspective of MMSE, NPI (S, CD), and DBD, the degree of dispersion was roughly equal, and concluded that there was no statistical difference between the two groups (p>0.1). It implied no massive difference between the two groups of participants in their cognition, the burden of nursing staff, and the degree of injury.

Linguistics Results

Sentence-ending particles

We recorded the conversations of the two groups of participants in the experiment word by word and counted the utterance of Japanese words and the sentence-ending particle "yo" and "ne" at the end of each sentence. And we used the T-test for the two groups of results (*Table 3*).

We found that the G1 group who directly communicated with the human coordinator in terms of utterance (Mean=1046.25, SD=876.5) is significantly higher than the G2 who teleoperated with the pepper robot (Mean=412.65, SD=525.8), with an effect size of Cohen's d=0.877 (r=0.401). For the sentence-ending particle "yo", there is no significant difference between the two invention methods (t(6)= 1.663, p>0.1). But the sentence-ending particle "ne" for the two groups was significantly different (t(6)= 2.313, p<0. 05), with an effect size of Cohen's d=0.837 (r=0.386).

Entropy of linguistic

We compared the Entropy of sessions during a more than 1-month experiment. The differences in the calculation results of 2-gram, 3-gram, and 4-gram are extracted respectively (*Table 2*), and the results by T-test detected the contrast of the results as follows (*Table 3*).

The analysis results of 2-gram, 3-gram, and 4-gram respectively, showed no significant difference between the two intervention methods (p>0.1). In the T-test result of 4-gram, the equal variance was not assumed, but no statistically significant difference was found between the traditional and robot groups (p>0.1). It can be considered that there is no difference between the two sets of results.

Emotion recognition results

From the first session to the last, the trend of changes in the emotional scores of the two groups of participants was relatively similar (Figure 6). The Emo-Rec point formula can conclude that the lower the score, the happier the participants. And the emotion detection results in the robot group showed a delighted trend.

Through the facial expression data of the two groups of participants with different intervention methods, we found that the differences between the two methods in various dimensions did not change whether carried out by robots or directly through humans. It implies that using robots instead of human labor under the premise of RGT can also satisfy the participants emotionally (*Figure 7*).

This experiment illustrates the change in the average point of the participants, at the beginning of 10.387, and next comes a decrease, ending at 8.531. It reflects that the participants' emotions had improved more positively during the process of RGT.

Discussion

The treatment or relief of PWD's BPSD generally consumes a lot of workforce and material resources, bringing work pressure to nursing staff and tremendous obstacles to PWD's personal life. In case of a labor shortage of nursing staff in the future, the use of robots to gradually save the workforce is an important topic (MHLW, 2020). With the help of robot studies in nursing facilities, we designed the intervention experiment in elderly facilities using Pepper robots to determine the remaining shortcomings of teleoperated robots in replacing the labor force to deal with BPSD.

We conducted an intervention experiment in a nursing home for more than one month in To-

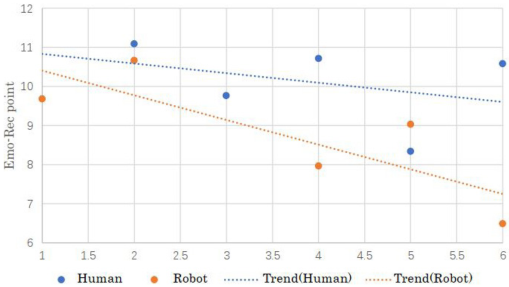

Figure 6. The trend of emotion point during RGT

kyo and then compared humanoid robots and human-coordinated RGT to help PWD alleviate the symptoms of BPSD. The MMSE scale was used widely in the study of dementia improvement, and it is also a more intuitive and convenient way to reflect participants' cognitive progress (Sugishita, Hemmi & Takeuchi, 2016). The use of the NPI-NH and DBD scales reflected from the perspective of the facility staff whether the two intervention experiments had reduced the nursing workload at the same level. Meanwhile, the two intervention methods have no difference in these scale results.

As an interventional experiment conducted for more than a month, the data reflected by the scales cannot provide more evidence, so we added the analysis methods of linguistics and facial recognition to offer more verification.

Linguistics analysis
The result was the same as our prediction: the robot group's utterance was significantly less than the human's. The teleoperation process will reduce the speed of human conversation due to hardware and network limitations, we also found the same result in the other teleoperated robot study (Kase et al., 2019). The redeveloped experimental system affected by the network quality should be considered within the results. The significant elements of network Quality of Service (QoS) defined by the TCP/IP (Transmission Control Protocol/ Internet Protocol.) protocol: packet loss, bit rate, throughput, transmission delay, availability, and jitter would affect the experiment results to a certain extent. Under the

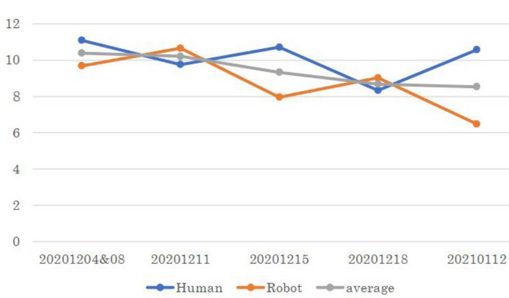

Figure 7. Emotion point during RGT

influence of the delay, the efficiency of the utterance between the participants and the operator will not share the same place in face-to-face therapy. Voice over IP technology (VoIP) relies on the TCP/IP protocol and has spread into all aspects of life, and the use of 802.11 wireless protocols still limits the response speed of the signal. In more application scenarios in the future, delay-free video/voice will gradually be introduced into daily life through 5G and other technologies. We believe that future network protocols and technologies can effectively improve QoS.

In Japanese, the sentence-ending particle "ne" is used to seek others' consent and empathize with others (Kinsui & Yukinori, 1998). In the comparison between the two groups, the sentence-ending particle "ne" was still at a different level ($p=0.047<0.05$). That may differ in the common ground between face-to-face and Human-Machine Interaction. Common ground is the knowledge, beliefs, and suppositions the participants believe they share about the activity. It is the foundation for all joint actions, making it essential to create the speaker's meaning and the addressee's understanding (Clark, 1996). When the dialogue partner is a robot, participants may think that robots have less common ground with humans. Therefore, they used "ne" less often.

However, there was no significant difference in sentence-ending particles "yo" ($p=0.108>0.05$). By mainly using the sentence-ending particle "yo", a speaker reminds the other party to pay attention to things they don't understand or share information with the other party (Kinsui & Yukinori, 1998). The result implies that both groups used the same soft tone, leading to a relatively similar communicative atmosphere.

Entropy is used to refer to concepts such as "randomness", "irregularity", and "ambiguity". It reflects the complexity of communication and the richness of the words carried (Shibuya, 2018). Complexity can be seen as uncertainty when the rules are complex and the expectations (or prediction) of the events that occur are difficult to predict (Ohmura, & Shibayama, 2016). The comparison shows that in the entropy result, there was no noticeable difference in the complexity and richness of the content between the two intervention methods.

Emotion recognition analysis
Although it is difficult for PWD to express their expressions, verifying emotions requires subjective results. SAR intervention tests usually use questionnaires to demonstrate teleoperating robots' feasibility and usability such as Face Scale and satisfaction surveys (Shibata & Wada, 2011).

Table 3. Comparison of linguistics results (n=8)

Variables		G1(n=4)M±SD	G2(n=4)M±SD	t	p
Utterance [a]		1046.25±876.49	412.647±525.80	2.434	.022*
Yo [a]		42.58±36.99	21.765±30.31	1.663	.108
Ne [a]		14.75±11.89	6.411±7.56	2.313	.029*
Entropy [a]	2-gram	3.20±0.02	3.22±0.03	-0.622	.548
	3-gram	3.25±0.02	3.28±0.03	-1.278	.230
	4-gram	3.24±0.05	3.28±0.01	-1.648	.130

Notes: G1 Traditional RGT group, G2 Robot GRT group, [a] Independent Samples T-test. *P<0.05

The trend graphs derived using the tool indicated that the effect of the robotic intervention improved the participants' mood more significantly. The result is in line with the theory of Heerink's, that robots will pleasant users at the first intervention stage by People Entertainment (PE). With the increase of time, people's confidence in the use of robots will also bring a better mood to users (Heerink, 2010). The significance of using the emotion recognition tool is mainly to solve the authorization restrictions and database problems encountered in using the previous software. The past face recognition software requires complicated authorization and use restrictions, which will inevitably cause researchers to extend the experiment period. There are also cases of the waste of experimental data due to the limitation of face recognition. Open source tools can flexibly adjust programming according to requirements and are more suitable for data collected in intervention research in terms of versatility. On the other hand, of the limitation of database matching, interventional study often selects experiments of different races in different regions, which requires the support of local people's databases to ensure recognition accuracy.

This method can improve accuracy by searching for a suitable database and thus providing more valuable research results. Trying to offer better verification ideas in the current situation that robots used to support PWDs are updated constantly and rapidly by finding better and more efficient face recognition ways or ideas.

In summary

In the process of conversation and feedback, we found that many participants discussed their shared experiences in their youth more actively. During the experiment, the operator failed to balance the time to give each participant to talk.

Regarding the selection of the operator of the robot intervention experiment, we found that older adults who can work are more suitable for this role. The reason is that vast age gaps often lead to difficulties in understanding and communication. Older adult patients with dementia recognize familiar matters better, but they need a slower acceptance process for strangers and new things.

Still, the older adult operators do not have the advantage of using robots or new technology products compared to the younger generation. With the modern network construction, nursing staff, and other staff can also control the robots remotely. The operator can conveniently access the therapy for PWDs at any time through the network and mobile communication technology. These change the form of labor and provide opportunities for the elderly to work.

As a solution to the labor shortage, the teleoperated robotic RGT method is currently feasible, reducing the pressure on caregivers and providing more labor opportunities for the elderly. They can become operators and need to read the user manual and then easily communicate with PWD according to the theme of the conversation.

Additional improvements and detailed research to identify and solve problems that current robots cannot perform in dementia nursing need to be conducted. We should conduct more studies with more possibilities of using robots to improve the labor shortage of nursing staff and recognize the tasks that robots cannot do and solve.

Limitations

By summarizing the intervention experiments, we identified some shortcomings. The first is the lack of sample size, and it is impossible to draw substantial conclusions through the comparison of the two groups. Looking forward to solving the problem of sample size in this type of research with the deepening of research in the future. Due to equipment and camera angle limitations, we could not fully generate a part of the data, which resulted in a lack of facial recognition samples. More convincing research may lead to obtaining by using more sophisticated and mature equipment in future studies.

Conclusion

The linguistic results for the teleoperated humanoid robot group were somewhat inferior. On the opposite, the PWD of the robot group in the mood was improved by a better trend. Conversational communication between humans and robots is not as smooth and natural as face-to-face communication. Improving this aspect would be an effective measure for robotic applications in cognitive care.

Acknowledgements

The authors would like to express their gratitude to Mr. Masato Yamazaki for his contribution in the development of Pepper's teleoperation system. This work was partially supported by JSPS KAKENHI [grant numbers 19K11395, 19H01577].

References

Ministry of health, labor and welfare. (2020, October). Reiwa 2nd Year White Paper on Health, Labor and Welfare. Ministry of Health, Labor and Welfare. https://www.mhlw.go.jp/stf/wp/hakusyo/kousei/19/index.html

Ministry of health, labor and welfare. (2020, October). Recent Changes in Socio-Economy and Household Trends. Ministry of Health, Labor and Welfare. https://www.mhlw.go.jp/content/000735866.pdf

Fuji keizai. (2022, March 8). 2022 Edition Current Status and Future Prospects of the Worldwide Robot-Related Market. Fuji Kei-Zai Group. https://www.fuji-keizai.co.jp/report/detail.html?code=162110822

Moyle, W., Jones, C. J., Murfield, J. E., Thalib, L., Beattie, E. R. a, Shum, D. K. h, O'dwyer, S. T., Mervin, M. cindy, & Draper, B. M. (2017). Use of a Robotic Seal as a Therapeutic Tool to Improve Dementia Symptoms: A Cluster-Randomized Controlled Trial. J m Med Dir Assoc, 18(9), 766–773. https://doi.org/10.1016/j.jamda.2017.03.018

Abdi J, Al-Hindawi A, Ng T, et al Scoping review on the use of socially assistive robot technology in elderly care BMJ Open 2018;8:e018815. doi: 10.1136/bmjopen-2017-018815

Liu, X. (2020). Social Assistant Humanoid Robots for Dementia of the Elderly: A Scoping Review. Journal of Electronic Research and Application, 4(6), 95–103. https://ojs.bbwpublisher.com/index.php/ssr/article/view/3317

Softbank. (2020). Pepper. Softbank Pepper. https://www.softbank.jp/robot/

Softbank. (2020). News of Pepper. Softbank Pepper. https://www.softbank.jp/sbnews/archive/category/Pepper

Carolis, B. D., Carofiglio, V., Grimandli, L., Macchiarulo, N., Palestra, G., & Pino, O. (2020). Using the Pepper Robot in Cognitive Stimulation Therapy for People with Mild Cognitive Impairment and Mild Dementia. The Thirteenth International Conference on Advances in Computer-Human Interactions, ACHI 2020, 452–457. https://www.thinkmind.org/articles/achi_2020_5_540_20203.pdf

Akutsu, Y. (2016). Pepper's Activity in the Field of Long-Term Care. The Japanese Journal of Medical Instrumentation, 86(4), 393–396. https://doi.org/https://doi.org/10.4286/jjmi.86.393.

Kase, H., Yamazaki, R., Zhu, W., & Nishio, S. (2019). Tele-Operated Android Robot Reminiscence Group Therapy and Human Coordinated RGT for Older Adults with Dementia: A Comparative Study. International Journal of Gerontology, Special Issue, S51–S55. https://doi.org/https://doi.org/10.6890/IJGE.201910/SP.0008

Moyle, W., Jones, C., Cooke, M., O'dwyer, S., Sung, B., & Drummond, S. (2014). Connecting the Person with Dementia and Family: A Feasibility Study of a Telepresence Robot. BMC Geriatrics, 14(7). https://doi.org/10.1186/1471-2318-14-7

Liu, X. (2021). Validation Research on the Application of Depth-Wise Separable Convolutional AI Facial Expression Recognition in Non-Pharmacological Treatment of BPSD. Journal of Clinical and Nursing Research, 5(4), 31–37. https://doi.org/10.26689/jcnr.v5i4.2325

Den, Y., Nakamura, J., Ogiso, T., & Oguta, H. (2008). A Proper Approach to Japanese Morphological Analysis: Dictionary, Model, and Evaluation. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA), 1019–1024. http://www.lrec-conf.org/proceedings/lrec2008/pdf/258_paper.pdf

Katagiri, Y. (2007). Dialogue Functions of Japanese Sentence-Final Particles 'Yo' and 'Ne.' Journal of Pragmatics, 39(7), 1313–1323. https://doi.org/10.1016/j.pragma.2007.02.013.

Hayano, K. (2017). When (Not) to Claim Epistemic Independence:The Use of Ne and Yone in Japanese Conversation. Journal of Pragmatics, 2(2), 163–193. https://doi.org/10.1558/eap.34740

Lodi, H., Saunders, C., Share-taylor, J., Cristianini, N., & Watkins, C. (2002). Text Classification Using String Kernels. Journal of Machine Learning Research, 2, 419–444. https://www.jmlr.org/papers/volume2/lodhi02a/lodhi02a.pdf.

Suzuki, M., Kuroiwa, R., Innami, K., Kobayashi, S., Shimizu, S., Minematsu, N., & Hirose, K. (2017). Accent Sandhi Estimation of Tokyo Dialect of Japanese Using Conditional Random Fields. IEICE Transactions on Information and Systems, E100.D(4), 655–661. https://doi.org/10.1587/transinf.2016AWI0004.

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and Entropy in Spoken Word Recognition: Effects of Age and Hearing Acuity. Exp Aging Res, 39(3), 235–253. https://doi.org/10.1080/0361073X.2013.779175.

Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shibuya, N. (2018, July 24). Entropy Demystified Is It a Disorder, Uncertainty or Surprise? Medium. https://naokishibuya.medium.com/demystifying-entropy-f2c3221e2550.

Clark, H. H. (1996). Using Language (9780511620539th ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Takanori, S., & Kazuyoshi, W. (2011). Introduction of Field Test on Robot Therapy by Seal Robot, PARO. Journal of the Robotics Society of Japan, 29(3), 246–249. https://doi.org/10.7210/jrsj.29.246

Ohmura, H., & Shibayama, T. (2016). Musical and Linguistic Structures Consisting of Pitch and Rhythm. The 30th Annual Conference of the Japanese Society for Artificial Intelligence. https://www.ai-gakkai.or.jp/jsai2016/webprogram/2016/pdf/398.pdf

Satoshi, K., & Takubo, Y. (1998). Danwa Kanri Riron Ni Motozuku 'Yo, Ne, Yone' No Kenkyuu [A Study of Yo, Ne, Yone in the Discourse Management Theory]. Onsei Ni Yoru Ningen to Kikai No Taiwa, 257–270.

Yoko, K., Toru, I., Akemi, K., & Maki, I. (2010). Reliability and Validity of an Observational Assessment Scale for Behavioral and Psychological Symptoms of Dementia (BPSD). Higher Brain Function Research, 30(4), 510–522. https://doi.org/10.2496/hbfr.30.510.

Sugishita, M., Hemmi, I., & Takeuchi, T. (2016). Reexamination of the Validity and Reliability of the Japanese Version of the Mini-Mental State Examination (MMSE-J). Japanese Journal of Cognitive Neuroscience, 18(3+4), 168–183. https://doi.org/10.11253/ninchishinkeikagaku.18.168.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model. International Journal of Social Robotics, 2(December), 361–375. https://doi.org/10.1007/s12369-010-0068-5.

Lorish, C. D., & Maisiak, R. (1986). The Face Scale: A Brief, Nonverbal Method for Assessing Patient Mood. Arthritis Rheum, 29(7), 906–909. https://doi.org/10.1002/art.1780290714.