早稲田大学審査学位論文
博士（人間科学）

Intervention Experiment of Reminiscence Group Therapy by Humanoid Robots for People with Dementia

ヒューマノイドロボットによる認知症グループ回想法の介入実験

2024 年 1 月

早稲田大学大学院　人間科学研究科
劉　相禹（リュウ　ソウウ）

LIU XIANGYU

研究指導担当教員：金　　群　教授

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Research Background

According to recently released population estimation data from Japan's Ministry of Health, Labour and Welfare, the number of older adults aged 65 and over in Japan increased by 320,000 from the previous year to 35.88 million, accounting for 28.4% of the total population. This proportion is expected to increase to 30.0% by 2025 and 35.3% by 2040 [1]. As the problem of aging has become more serious, it has had multiple negative effects on Japan's economy and society. The burden of pensions and medical care continues to increase, and the finances of the Japanese government are under enormous pressure.

As the demand for labor and human resources increases with the aging of society, there is a need for care services. According to statistics from the Ministry of Health, Labour, and Welfare, approximately 8% to 10% of people with dementia (PWD) (including vascular dementia and Alzheimer's disease) are aged 65 years and older [2]. All older adults born during the first postwar baby boom will be over 75 years of age by 2025, and there will be a shortage of approximately 270,000 care workers. Institutions estimate that there will be a shortage of 680,000 care workers by 2035 [3]. A survey of labor market trends (2019) also showed that approximately 40% of medical and nursing homes reported staff shortages (including full-time and part-time care facilities) [4].

Faced with these problems, the Japanese government has attached substantial importance to the development of intelligent technology (IT) for older adult care and promoted it as an economic industry. With the pace of IT development, Japan is making full use of innovative technology to provide comprehensive care for older adults. The market for medical and nursing robots is expected to reach JPY 534 billion by 2025, an increase of 2.21 times compared to 2020 [5]. Thus, the future of the robotic economy is a promising growth situation. The use of robots in nursing is an essential solution to the current labor shortages.

Since machines have replaced some human resources, various attempts have been made to use industrial robots as robotic arms in the medical industry. After 2000, with the development of technical strength, more machines, such as early telerobots, were being used to augment human resources. In addition, teleoperated devices have begun to play an important role in medical and nursing care [6].

The PARO was designed in 1993 and launched in 2001. It was the pioneer of research on the care of older adults with dementia and the alleviation of symptoms, such as Behavioral and Psychological Symptoms of Dementia (BPSD), by robots [7]. Subsequently, the concept of socially assistive robots (SAR) was introduced. Robots for the care of older adults and those with dementia have gradually been developed and are widely used in facilities worldwide [8]. With the development of industries such as semiconductor and manufacturing, socially assistive humanoid robots (SAHRs) have gradually been used in daily life [9]. Compared with previous pet-type robots, the SAHR can achieve more symptomatic applications for specific manifestations, such as BPSD. For example, by alleviating the peripheral symptoms of BPSD through human-computer interaction, PWD can return to society through counseling and communication [10].

1.2 Research Purpose

BPSD is a collective term for the manifestations in PWD due to cognitive impairment. The core BPSD, including memory problems and disorientation due to brain lesions, lead to poor comprehension and language skills. Peripheral symptoms also cause problems such as physical behavior, wandering, and depression. BPSD have a complex presentation that place a significant burden on care givers [11]. Effective interventions should address both physiological and psychosocial factors. Also, rather than the "physiological factors" approach, the "psychosocial factors" strategy has been widely used to help PWD respond to changes in the environment and social relationships [12]. The robot is expected to fulfill its role with sufficient effectiveness when involved in the care and non-pharmacological treatment of PWD [13].

The Reminiscence Group Therapy (RGT) method is currently more feasible from the perspective of labor savings [14]. Traditionally, non-pharmacological countermeasures for BPSD relied mainly on professional psychophysical therapists for treatment. The treatment process cannot be separated from that of a large number of skilled caregivers [11,13-14]. Now that SAR robots are more mature, we need a framework to evaluate whether SAHRs can effectively treat BPSD in PWD to ensure that their effects can be effectively assessed in interventional trials [15].

During COVID-19, traditional working structures have been challenged. Due to epidemic prevention considerations, some older adult care facilities have restricted entry and exit rules, which have also put pressure on PWD and nursing staff. First, family members of older adults in institutions can only communicate via telephone and video, which does not provide emotional relief or pressure. In addition, the range of care staff activities has been compressed, resulting in higher work pressure. The relationship between this stress and BPSD of PWD has been pointed out in some studies, and researchers are called on to pay attention to this relationship and seek improvement.

Based on the summarized literature, we identified the issues that must be considered when using SAHRs to alleviate BPSD and improve it in field experiments. Finally, this thesis focuses on improving these issues and demonstrating the effectiveness of the SAHR in replacing the workforce.

1.3 Organization of the Thesis

This thesis is divided into five chapters, as shown in Figure 1-1. Chapter 1 examines the practical issues resulting from the pandemic and recent research on robotic dementia nursing. This chapter aims to establish a clear position for this thesis and outlines the direction and research methods within a limited scope.

Chapter 2 is a scoping review that analyzes and summarizes existing SAHR-related research on dementia and utilizes the SAR evaluation framework to better determine the roles of SAHR in the mitigation of BPSD. It reveals the current research methods that are suitable for utilizing SAHR to assist PWD. The scoping review conducted a search for articles published around May 2020 and found that SAHRs are needed for performance and future directions in emotional recovery, social facilitation, and cognitive improvement to enhance the relief of BPSD symptoms in PWD. The effectiveness of the intervention-based dialogue experiment was confirmed.

We selected Pepper for our intervention experiment because it has already been integrated into mass-produced robot models. In addition, our scoping review indicates that major articles anticipate the integration of automatic emotion recognition. By leveraging Python's deep-learning artificial intelligence (AI) and open-source programming tools, we developed an Emo-Rec Application to automatically detect the emotions of PWD.

In Chapter 3, we assess the usability of the tool with the aid of three staff members and experts experienced in the care of older PWD. They evaluated the expressions of PWD during the intervention experiment and compared them with the results of the Emo-Rec App. Consistent results were observed between the two evaluations.

The Emo-Rec App has the potential to recognize facial emotions efficiently and accurately in PWD and serves as a theoretical foundation for future use. Our validation of the Emo-Rec App using experimental data from a nursing home in Tokorozawa revealed how teleoperated RGT can aid mute older PWD in emotional improvement.

In Chapter 4, we conducted an interventional experiment lasting more than one month utilizing two methods of RGT therapy for PWD: face-to-face and a teleoperated humanoid robot. We adopted emotion recognition and linguistics to investigate whether robots can replace humans. These positive results suggest that robots can feasibly mitigate the BPSD of PWD, to a certain extent.

Chapter 5 discusses the results of various experiments. It also explores the expected impacts of future technological developments in the field. One key discovery is the potential application of humanoid robots in facilities, whereas another crucial contribution is the quantification of improvement in PWD with regard to the emotional dimension.

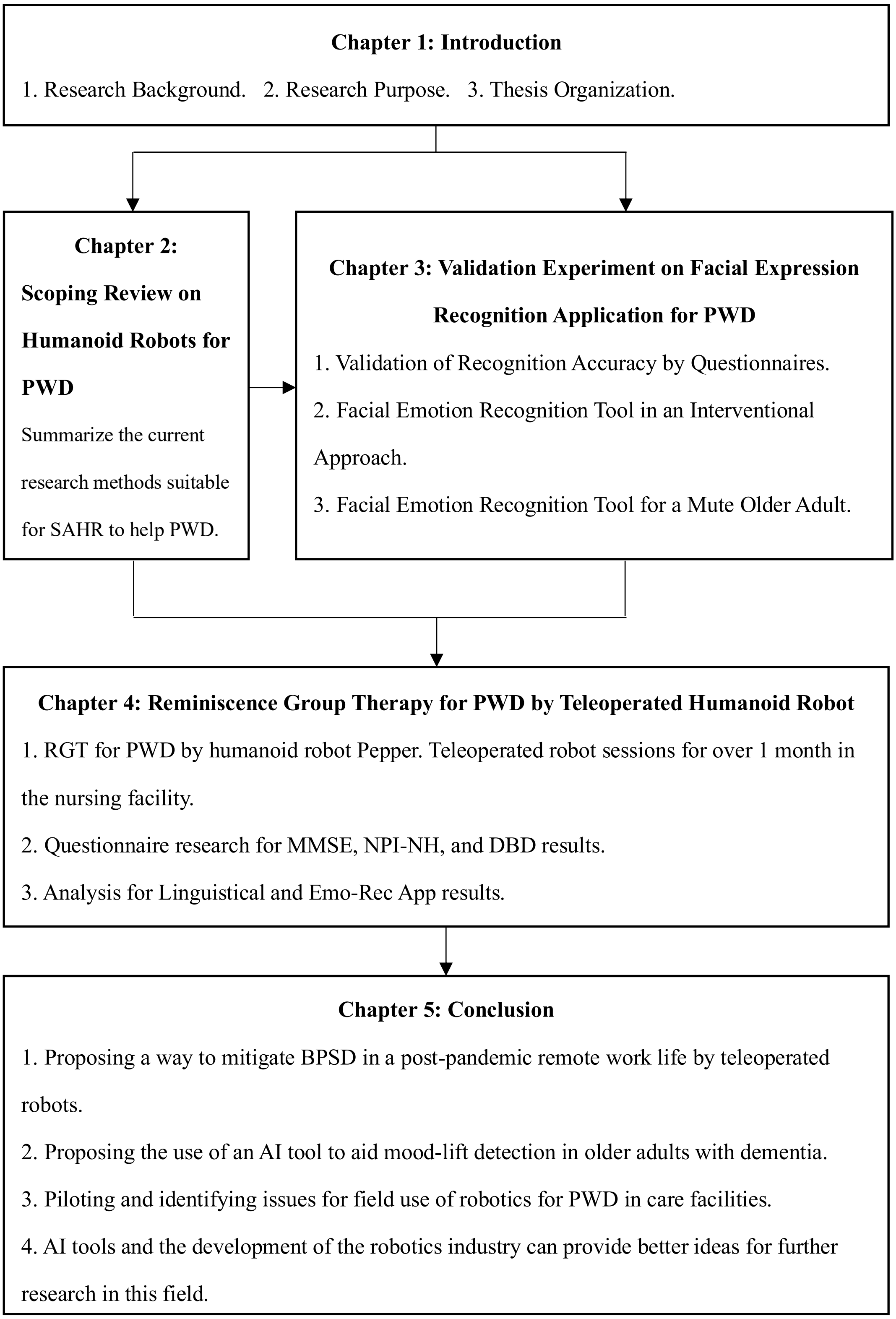


Figure 1-1. Thesis Organization

Chapter 2 Scoping Review on Humanoid Robots for PWD

2.1 Background and Purpose

Recently, robots have become capable of performing tasks previously performed by humans and are increasingly utilized in medical services. Today, society is faced with a workforce shortage and numerous scholars and researchers are attempting to develop novel solutions to alleviate the symptoms of dementia by utilizing robotics and AI [13,15].

Conversational robots have gained wide attention in the treatment of cognitive impairment. Instead of companion robots, conversational robots provide care assistance by fulfilling social participation through dialogue and actions, thereby effectively alleviating the symptoms of dementia. [16]. Conversational robots are functionally opposed to pet-type robots and mainly refer to robots capable of realizing conversational functions. A SAHR is a relatively new concept in conversational robots [10]. A significant amount of medical and pathological research has been conducted and documented on pet robots such as PARO [8]. However, research on humanoid robots is limited.

In this study, we explored a SAR framework to define and discuss humanoid robots. Currently, pet robots such as PARO, Aibo, and Pleo have been studied, and useful models have been established to aid in comprehending the role of robots in older adult care [8, 15, 17]. The SAR framework has five roles: affective therapy, cognitive training, social facilitation, companionship, and physiological treatment. We utilized the SAR model to assess the SAHR research performance.

The purpose of this scoping review is to examine the effectiveness of SAHR in the context of the SAR model. Based on previous research findings [15], humanoid robots are more effective than pet-type robots in cognitive training. Therefore, our focus was on the use of humanoid robots to train cognitive abilities through dialogue and other interactions that are not present in pet-type robots. These interactions provide additional training effects within the proposed framework.

2.2 Methods

We conducted a scoping review using the Joanna Briggs Institute method (2015 version) and the PCC mnemonic (P = Population, C = Concept, and C = Context). Our research aimed to assess the impact of humanoid robots on individuals with dementia. Given the increasing use of humanoid robots is a growing trend, we sought to develop a comprehensive framework for evaluating future robots.

The study population included individuals with dementia and their caregivers. The concept of this framework requires further development by researchers and the context relates to the use of SAHRs. These elements correspond to the PCC mnemonics.

2.2.1 Paper Search and Selection Method

In this study, we searched for publications on older adult care robots published from 2009 to 2019. The search was conducted using a comprehensive set of databases, including NDL, IEEE Xplore, CiNii, Web of Science, and Google Scholar. The keywords used in the search were "Humanoid robot," "Elderly care," and "Dementia," as listed in Table 2-1. After analyzing the available literature, we narrowed the scope to cover the period from 2000 to 2020. The search primarily targeted the period in March 2019. The most recent search was conducted in May, 2020.

Table 2-1. Keywords

Humanoid Robot	**Elder Care**	**Dementia**
“Social assistant Humanoid robot” OR “Social assistant robot” OR “Nursing robot” OR “Humanoid” OR “Care robot” OR “Humanoid Robot”	“Elderly care” OR “Hospital care” OR “Facility care” OR “Elder adult care” OR “Elderly adult care” OR “Eldercare”	“MCI” OR “PWD” OR “OAWD” OR “Cognitive” OR “Alzheimer's” OR “Dementia”

Duplicate articles were excluded during the initial screening. The discussion was restricted to humanoid robots that were either self-designed or mass-produced (excluding pet robots) and met specific inclusion criteria. Our focus was primarily on humanoid robots that underwent long-term testing in an experimental environment designed by the facility. These steps were used to filter and integrate the final results as Figure 2-1 shown.

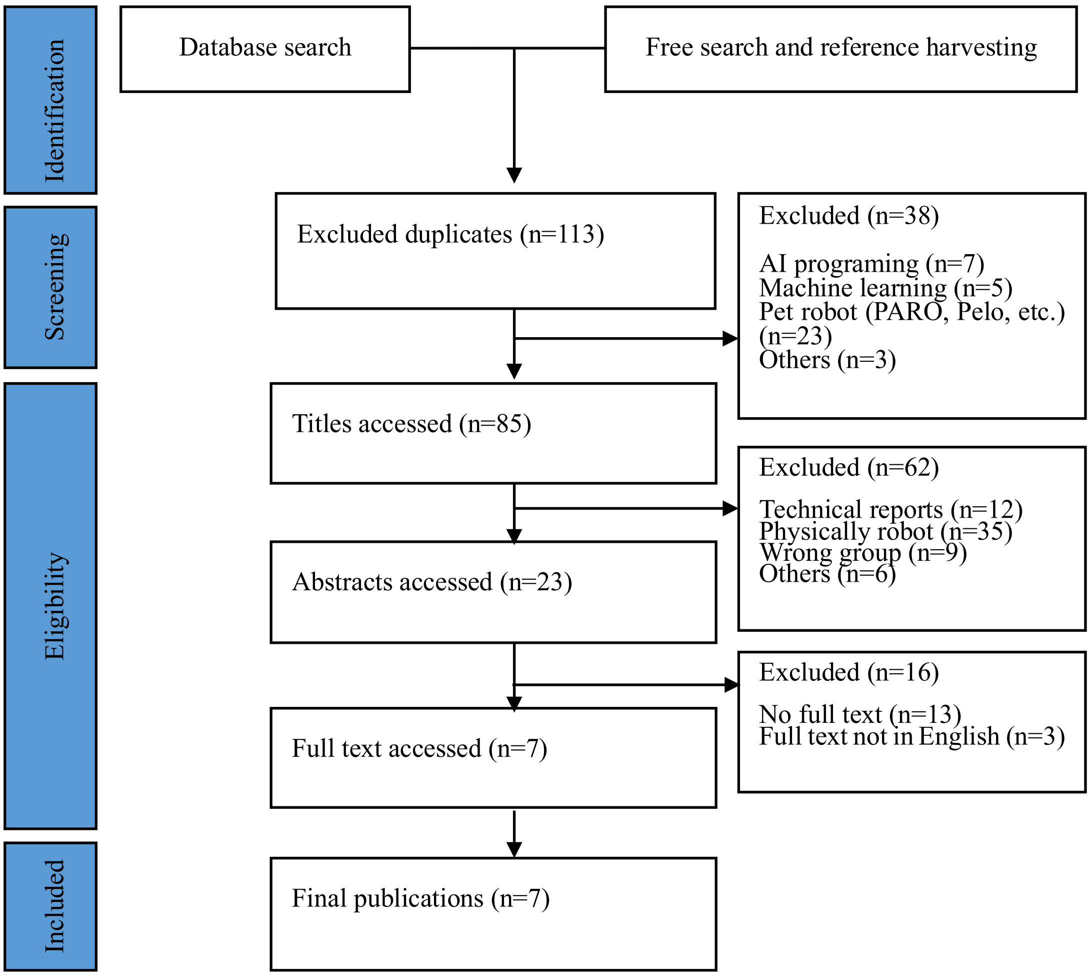


Figure 2-1. Flowchart of the Identification and Selection Process

2.3 Results

2.3.1 Characteristics of the Literature

A total of 137 articles were found and seven papers [18-24] were retrieved. For comparison and discussion, six humanoid robots were used in the seven studies. After excluding unspecified sex characteristics, female participants accounted for 94.45% of the total. These studies were conducted in Europe, Oceania, and North America across five different countries. Except for two studies [20,23], the duration of the intervention did not exceed two days.

Three studies, with more than 12 subjects each, focused on mild cognitive impairment (MCI) [18-19, 24]. Most studies were conducted in care facilities. Nursing staff surveys and follow-up are essential components of all studies to ensure proper attention for PWD. The inclusion or exclusion of a Graphical User Interface (GUI) can affect the quantification and evaluation of results. The features of the robots are listed in Table 2-2, and their appearances are shown in Figure 2-2. An overall summary of the seven studies is presented in Table 2-3.

Table 2-2. Characteristics of the Humanoid Robots

	GUI	**Wireless Network**	**Voice Call**	**Video Call**	**Mass-Produced**
Brain 2.1	-	-	+	-	-
Telenoid	-	+	+	-	+
ED	+	-	-	-	-
Giraff	+	+	+	+	+
Kompai	+	+	+	+	+
Healthbots	+	+	+	+	+

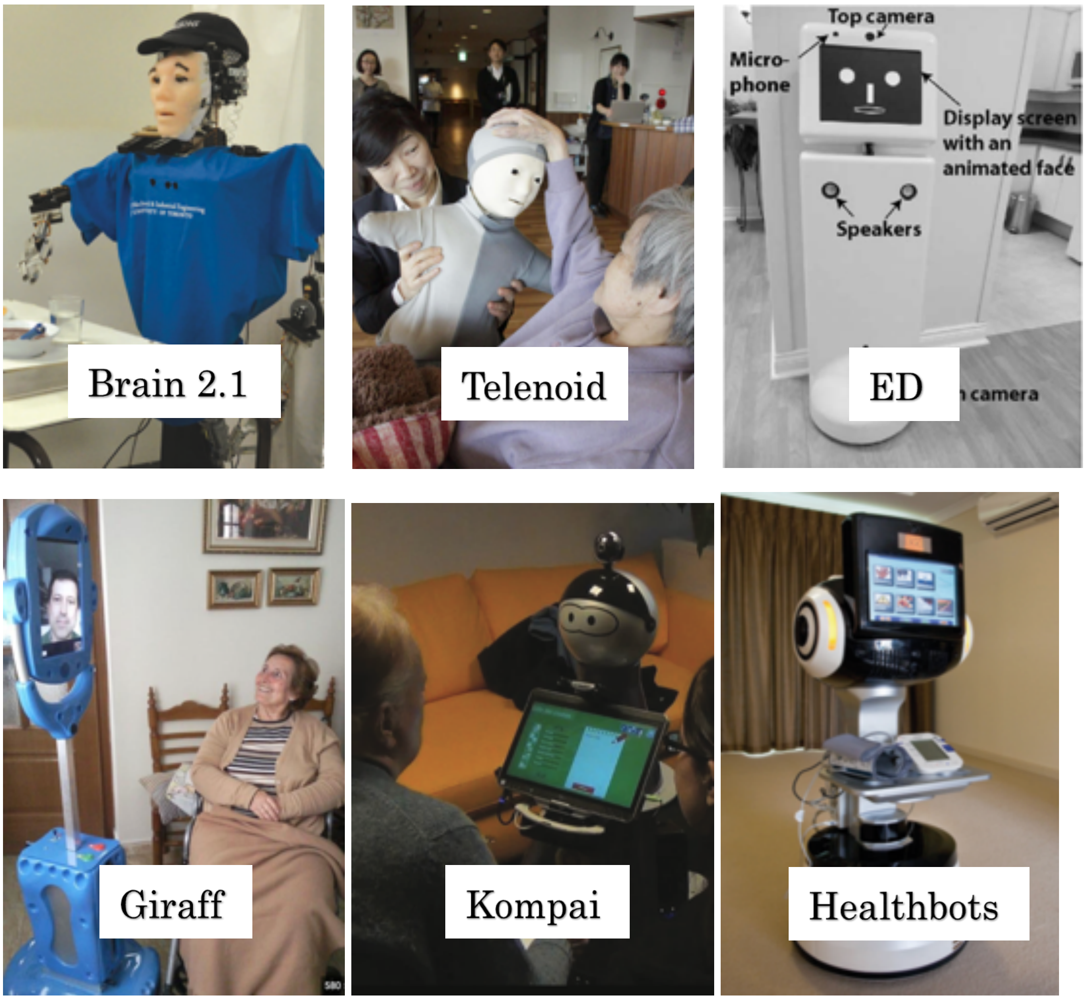


Figure 2-2. Appearance of Humanoid Robots

Table 2-3. Characteristics of the Selected Literature

	Ref.	Robot	Intervention/study design	Setting	Participants	Duration	Measure	Outcome
Brian 2.1 A Socially Assistive Robot for the Elderly and Cognitively Impaired	McColl (Canada)	Brian 2.1	The research measures the involvement of card games and time and interaction to measure robot feeding, acceptance, and attitude.	long-term care facility	MCI (n=40)	Two days	Questionnaire DNSS nonparametric Mann–Whitney test Cronbach's alpha (a) Spearman's P	1. The participants had a positive experience with the robot, which influenced their motivation to use it again. 2. Regardless of their computer experience, both male and female participants found the robot easy to use. 3. HRI study presented promise for using a human-like robot for cognitive interventions and motivated further development and long-term testing of the robot.
Teleoperated android as an embodied communication medium: A case study with demented elderlies in a care facility	Yamazaki (Denmark)	Telenoid	Talk to a Telenoid for 20 minutes and conducted a qualitative method with the dialogue recording data of the elderly.	long-term care facility	PWD (n=30)	20 minutes	Questionnaire CIRCA, Networked Interaction Therapy, ECT	1. The older adults become more willing to converse with the Telenoid, and the operator can easily communicate with them by responding to their positive reactions based on their growing affection. 2. It slowly elicited their spontaneity to communicate with others, whereas the symptoms of dementia include apathy.
Collaboration of an assistive robot and older adults with dementia	Begum (Canada)	Ed	Behavioral observations, semi-structured interviews, and questionnaires facilitated the collection of in-depth and diverse data to inform the design of an assistive robot for OAwDs.	iDAPT HomeLab	PWD (n=5) MMSE scores 9, 24, 25, 25, 18	2.5 hours	Questionnaire MMSE, semi-structured interview	1. Hard to use the information provided by the human partner with limited perception and cognition. 2. Most of the OAwDs showed some level of situational awareness. 3. Failure to develop a correct mental model of the robot. 4. Robot for HRI needs to offer enough sophistication.
Connecting the person with dementia and family: a feasibility study of a telepresence robot	Moyle (Australia)	Giraff	Five people with dementia and their family members participated in a discussion via the Giraff robot a minimum of six times over six weeks.	long-term care facility (Queensland, Australia)	PWD (n=5) five residents, six families, and seven staff	six times over six weeks	Questionnaire The semi-structured interviews OERS Noldus ObserverXT 11.5 program	1. Positive emotions tended to taper off in the middle of the call. 2. Dyads used the opportunity for visual cues 7.26 times (SD=4.59) in each call session to engage with the family member. 3. Participants perceived Giraff as a positive and therapeutic option to engage people with dementia with their family members.
Robot services for elderly with cognitive impairment: Testing usability of graphical user interfaces	Granata (France)	Kompai	A group of 11 older people with Mild Cognitive Impairment and 11 cognitively healthy older individuals took part in this study. Performance measures (task completion time and the number of errors) were collected.	Broca Hospital	MCI (n=11) and cognitively healthy older adults (n=11)	One session	Wilcoxon signed-rank Kruskal-Wallis chi-squared tests Observer R XT	1. Participants in the MCI group (M = 375.16 s) were slower than HC (M = 355.76 s) at completing tasks. 2. Results showed that the speed of performance improved significantly in both groups when adding a product to the list for the second time (MCI: z = 2.38, p = 0.009; HC: z = 2.2, p = 0.01). 3. These findings suggest that two trials were enough to see initial speed differences between age groups disappear.
Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the living lab setting	Wu (France)	Kompai	Participants interacted with an assistive robot in the Living Lab once a week for four weeks. After being shown how to use the robot, participants performed tasks to simulate robot use in everyday life.	LUSAGE Gerontechnology Living Lab (located in a building of the Broca Hospital, Paris, France)	MCI (n=6) and CIH (n=5)	once a week for four weeks	Questionnaire Semi-structured interview Wilcoxon matched-pairs test	1. In the CIH group, there was a tendency (P=0.07) toward a decrease in scores in the "attitudes toward robots" dimension. 2. Completion time increased significantly in the MCI group (P=0.028). Errors (P=0.27) and help (P=0.67) in the MCI group and completion time (P=0.50), errors (P=0.79), and help (P=0.59) in the CIH subjects remained unchanged.
Improved robot attitudes and emotions at a retirement home after meeting a robot	Stafford (New Zealand)	Healthbots	Participants were given two questionnaires, one before and one after interacting with the robot. After using the robot, participants rated the overall quality of the robot interaction.	Selwyn Village	Residents (n=32) and staff (n=21)	30 minutes	Questionnaire MANOVA PANAS, RAS	1. Three of the eight pre/post-interaction attitudes and emotions were significantly correlated with robot rating. Pre-interaction attitudes and feelings are essential for robot acceptance. 2. After meeting the robot, there were significant improvements in the attitude of the participants towards robots (F (1, 27) = 25.04, partial eta squared =.48) as well as decreases in negative affect (F (1, 27) = 4.30, partial eta squared =.14).

2.3.2 The SAR Framework

(1) Affective Therapy

An analysis of six studies involving 113 participants (149 in total) in the field of humanoid robots indicated that conducting a series of experimental studies on such robots could have a positive impact on the overall mood of individuals [18-21,23-24]. All six studies identified significant positive outcomes from the experiment, such as a decrease in depression score, an increase in agitation score, and Quality of Life (QoL) score [18-21,23-24].

Derek et al. determined that 82% of the participants displayed a smile when the robot gave an entertaining performance. Of the seven participants who became distracted at least once during the interaction, 57% responded verbally to the emotions of the robot and returned for the experiment. The remaining 43% refocused on robots and activities [18].

Although individuals with disabilities have difficulty sustaining verbal and nonverbal interactions with a Telenoid, Yamazaki et al. found that they gradually engendered spontaneous communication with others. The symptoms of dementia, including indifference, were alleviated to varying degrees [19].

According to Moyle et al., residents tended to display the most positive emotions at the start of a call when locating their families on a Giraff screen. The position of the positive emotional call gradually diminished during the call, and the emotional response of the resident depended greatly on the nature and content of the conversation [21].

Additionally, a study conducted by Stafford et al. demonstrated that over 80% of participants were open to receiving robot assistance and companionship [24].

(2) Cognitive Training

Four intervention methods were identified for six robot types. Table 2-4 demonstrates that dialogue is the most significant intervention method for humanoid robots in cognitive training. In addition, assigning tasks, games, and diet assistance have been mentioned in cognitive training.

Table 2-4. Intervention Method Summary

Robot	Assign task	Dialogue	Game	Diet Assistance
Brian 2.1	+	+	+	+
Telenoid（R1）	-	+	-	-
ED	+	-	-	-
Giraff	-	+	+	-
Kompai	+	+	+	-
Healthbots	-	+	-	-

Five studies involving 103 patients with MCI or PWD (149 in total) yielded positive results. In the Brain2.1, Ed, Kompai, and Healthbots experiments, it was observed that participants required the assistance of a robot to complete the task in the experimental project [18, 20, 22-24]. For instance, the participants engaged in various activities that reflected aspects of contemporary social life, such as playing card games (Brain 2.1), brewing tea (Ed), using robots for assistance, engaging in online communication, and virtual shopping (Kompai). Two of the three studies compared the results with those of healthy adults and found that individuals with MCI and PWD exhibited improved task completion times and accuracy rates in post-learning evaluations. A summary of cognitive training studies is presented in Table 2-5.

Table 2-5. Cognitive Training Summary

Intervention method	Studies	Feasibility	Quantity	Completion
Assign Task	2	+	+	+
Dialogue	4	+	+	+
Game	3	+	+	-
Diet Assistance	1	-	-	+

Granata et al. (2021) found no significant difference in the time required for young people and those with MCI to complete task operations, even with the same learning cost status, under reasonably configured GUI settings [22]. However, the MCI group took longer to perform complex tasks because of age-related and physiological conditions. However, older individuals with MCI or dementia can still benefit from intervention tasks performed by a robot with a GUI.

In the cognitive training comparison experiments, researchers obtained improved results by comparing the effectiveness of the two situations, with and without robot assistance. In addition, the questionnaire related to the Brain 2.1 experiment received a "recognized usefulness" score of 4.37 (out of 5.0, SD=0.96) [18].

(3) Social Promotion

In five studies, the researchers assessed the effectiveness of humanoid robots in promoting social interactions among 101 participants (149 in total), with positive outcomes [18-21,24]. According to surveys conducted on social presence of Brain 2.1, participants reported feeling as if they were conversing with real individuals and noted that the robots maintained eye contact during interactions [18]. The project achieved a rating of 3.46 out of 5.0 (SD=1.46) in comparison to Kompai's survey, which received a score of 3.17 out of 4.0 (SD=1.19) in the section regarding "social impact." Yamazaki et al. determined that the Telenoid has the potential to be an advantageous tool for the comprehension of others, particularly in dementia care [19]. Nonverbal interactions with the Telenoid demonstrated the ability to instigate impromptu communication among older adults, who typically displayed limited interest in conversing with others or dolls. According to recent studies, Telenoid provided a temporary "conversational stepping stone" to improve QoL and regain conversational and social confidence [25].

These findings suggest that robots can effectively address loneliness and social isolation. However, further research is necessary to understand the long-term effects of robot usage and address any potential concerns or challenges related to their utilization.

(4) Companion

Three studies involving 78 participants (149 in total) were conducted to assess the use of robots to combat loneliness and social isolation [21-23]. The results demonstrated a favorable association between the use of robots and enhanced social integration. Notably, two of the studies, Giraff and Kompai, involved experimental durations exceeding one month. However, the other studies discussed in the experiment only followed the interventional approach for a maximum of four months. Long-term survey experiments may be necessary to prevent data errors caused by the Hawthorne effect.

Yamazaki et al. highlighted the possibility of older adults rejecting and feeling apprehensive about the use of robots as companions [19]. The need to define the roles of robots in companionship is emphasized by the challenge of quantifying their dimensionality.

Surveys in these studies displayed the degree of anthropomorphism attributed to the robot by the participants, their level of confidence in the robot, and its potential to serve as a long-term companion [26]. However, it is important to recognize that short-term experiments may be influenced by the novelty of the robot, potentially resulting in biased results. Therefore, these findings may not accurately reflect the long-term impact of the use of humanoid robots, an area that has not been fully explored in these studies.

(5) Physiological Treatment

Previous research on robot evaluation has commonly used the PARO and Nao series robots as reference points. Pet-type robots have generated adequate data to demonstrate the “physiological indicators” framework within the SAR framework. However, further research is required to effectively determine SAHR under this framework. An accurate and comprehensive framework is necessary to compare and assess the benefits, drawbacks, and effects of robots. Alternatively, conducting additional randomized controlled trials (RCTs) related to SAHR is crucial for enhancing physiological index data and preventing misinterpretations among readers and researchers.

2.4 Discussion

The reviewed articles exclusively considered humanoid robots, highlighting the challenges researchers face compared with traditional companion pet robots. Of the six robots described in the literature, only three are mass-produced, whereas the other two are custom-made with low-volume production [18,20]. The primary objective is to enhance robotic products through exploratory experiments or experimental research utilizing these robots. In future studies, the implementation of advanced mass-produced robots will efficiently address this situation.

The SAR framework used in this study was only suitable for some of the seven studies. Unfairness lies in the comparison between a robot with a GUI and a communication- and emotion-providing robot without a GUI. It is complex to assess the advantages of a robot with specific functions. Currently, a humanoid robot is defined based on its physical appearance and user feedback. However, the design intent must align with the application scenario [22].

For instance, feature phones and smartphones share the same mobile phone category, but differ significantly under the same framework. Hence, there is a need for a framework that can precisely evaluate the advantages and disadvantages of robots and their effectiveness and reflect them in an objective manner.

Cultural variations among the studies merit attention, as only one study originated in Japan, whereas the others were conducted in Europe and North America. The influence of language on the emotions of both the robot and the operator must be considered.

We learned that multiple SAHRs with GUI have built-in multi-language support. In remotely controlled robots, the operator can communicate in the same language. Additionally, some of the studies mentioned in the Discussion involved a significant proportion of female participants (almost 95%). It is essential that the conclusions presented in this article avoid potential biases resulting from such factors.

2.5 Summary

Currently, SAR provides an evaluation framework for SAHR; however, limitations and deficiencies still exist. The SAR framework proposed by Abdi et al. also includes a standard for measuring physiological indicators. However, we have not found any research that can quantify this indicator in current research on humanoid robots.

The increased use of humanoid robots in the future is likely to yield more comprehensive data and enable more objective evaluations. Conducting studies on humanoid robots akin to RCT for pet robots is critical for promoting 'advancements in the field.

The development of communication technology also directly affects the applications of humanoid robots. The 5G/Wifi6 wireless technology allows voice or video interaction with reduced latency and clearer images. Humanoid robots capable of displaying remote operator actions can instantly utilize more body language to enhance communication complexity. Advancements in robot technology can also cater to on-site requirements such as social interaction, companionship, and human-like abilities.

Overall, there is a current issue with the evaluation framework for the SAHR meeting current needs. However, we anticipate substantial advancements in humanoid robots through objective evaluation and technological progress.

Chapter 3 Validation Experiment on Facial Expression Recognition Application for PWD

3.1. Background and Purpose

3.1.1 The State of Emotion Recognition in the Study of Dementia

Emotion refers to the way individuals assess whether external objects or events meet their needs. Current research on the interaction between emotions and the brain suggests that emotions are the brain's forecasts of external reactions. It is an influential and reflective instrument for assessing brain activity in other dimensions. According to Richard et al. [27], facial expressions are directly processed by the limbic system, making them a valuable way to measure and observe individuals' emotions and needs. Alongside spoken language, facial expressions are a nonverbal way for individuals to convey their emotions. In studies on dementia, it is imperative to focus on the psychological activities and emotional needs of individuals with dementia as reflected in their facial expressions [12].

Notable clinical presentations of dementia or BPSD are a lack of emotional expression, an increase in negative emotions, and inconsistent emotional states. Observing a decrease in BPSD through emotional assessments is a commonly used and appropriate testing method for non-pharmacological therapy [28]. Understanding and recognizing the emotional requirements of people with dementia are crucial, and facial expressions can be used to directly observe and assess their emotional states. AI image recognition systems can offer valuable insights into the emotional performance of older individuals with dementia, ultimately leading to improved non-pharmacological therapeutic interventions [21].

3.1.2 Current Challenges of Facial Emotion Recognition

Currently, three primary methods are used for measuring emotions: emotional facial action coding systems (EMFACS), electromyography (EMG), and automatic face recognition. According to Karsten et al., the EMFACS could be more effective, which involves analyzing fixed images individually based on the specifications of the system [29].

Previous research has demonstrated that the EMFACS can accurately measure emotional expression in non-experts; however, image scoring requires highly trained professionals to maintain objectivity. In addition, facial coding must be scored manually, and the EMG system entails technical challenges and limitations in experimental settings, such as positioning sensors to track facial muscle movement. Using this system to examine facial expressions in everyday social scenarios is not feasible.

Karsten et al. contended that automatic facial recognition can potentially replace pre-existing methods [29]. Automatic facial recognition technology can identify and analyze facial expressions more efficiently and accurately. It is more convenient and applicable in daily life because it does not require sensors or equipment to be attached to an individual's face. Although the EMFACS and EMG methods have been used to quantify facial expressions and accurately measure emotions, they suffer from technical complexity and experimental limitations. Automatic facial recognition technology can potentially replace these methods, providing a more convenient and accurate approach for identifying and analyzing facial expressions in daily life.

3.1.3 Improve Efficiency Through Depth-wise Separable Convolution

In 2017, Google implemented a novel computing architecture in MobileNet for depth-wise separable convolution in convolutional neural networks. The aim was to reduce the computational burden of the CNN (Convolutional Neural Network) model [30]. This technique fundamentally dissects the standard convolution calculation method into depth-wise and point-wise convolutions, thereby preventing the output structure from being affected by the reduced number of calculations [31]. By utilizing fewer system resources, AI deep learning tasks that require significant computing power can now adapt to more lightweight scenarios, increasing their versatility and applicability to a wider range of needs [32-35]. In 2017, Arriaga et al. successfully incorporated real-time convolutional neural networks for emotion recognition in robots, further supporting the use of facial emotion recognition in robotic interventional experiments [36].

3.2 Emo-Rec Application and Face Scale

Using open-source tools, we can objectively quantify and analyze mood fluctuations among PWD in social settings, as depicted in Figure 3-1. The resulting detection output categorizes expressions into seven emotions: neutral, anger, contempt, disgust, fear, happiness, and sadness. This study used a dataset composed of 28,709 training samples, 3,859 validation datasets, and 3,859 test samples. The dataset contained 35,887 images classified into seven categories: angry, disgusted, fearful, happy, sad, surprised, and neutral. Each image had a resolution of 48x48 pixels. A significant portion of the images in this dataset present horizontal and vertical rotations, and many are partially obstructed by occlusions such as hands, hair, and scarves. Our training model was integrated into the FER2013 dataset, which achieved a classification accuracy of 86%. These tools were used to complete the initial construction of Emo-Rec

Application. We applied this tool to the experimental data of patients with BPSD and dementia to assess its viability.

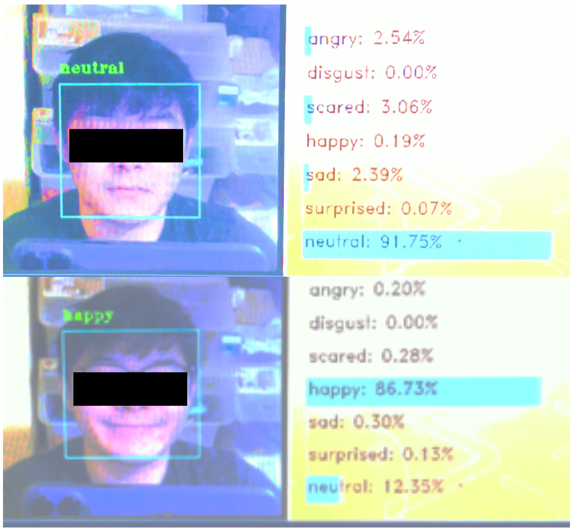


Figure 3-1. The Emo-Rec Application

The Emo-Rec Application uses a camera to identify faces and searches a database for matches. Emotions in patients with dementia were evaluated by this application using objective data, resulting in higher efficiency. The Face Scale is a tool frequently used for patients with emotional impairments [37]. Participants were instructed to select from a range of 1–20 expressions and provide feedback to the researchers as Figure 3-2 showed.

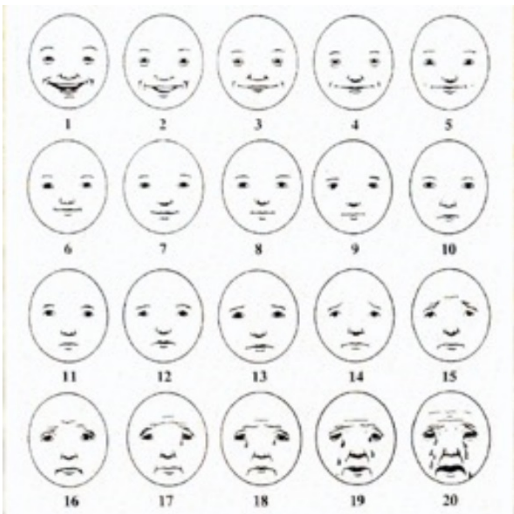


Figure 3-2. The Face Scale

This study utilized a 1–20 scale to enable effective communication with personnel and for data comparison. The gathered facial expression data can be transformed into a 1–20 scale for analysis by applying the weighted formula outlined in Formula 3-1.

Emo-Rec point=(anger*10) + (disgust*10) + (scare*10) + (happy*1) + (sad*20) + (surprise*10) +(neutral*10)

(3-1)

The output of the Emo-Rec App was presented as a percentage-matching degree. The app ratio ensured that the total emotional ratio was 100%. Using this formula and the app ratio, we created a range of 1–20. For instance, happy100% equates to 1 and sad 100% equates to 20. This formula was used to convert the output result into the more commonly used Face Scale. This survey instrument aims to gather genuine sentiments of PWD via a questionnaire as Figure 3-2 showed.

The Face Scale is typically used to convey emotions and sensations in Japanese patients with dementia. This instrument was used to enhance communication and collaboration with facility personnel. It is possible to visualize 0 on the Emo-Rec Point scale as equivalent to 10, expanding the total range from -10 to 10. To compare the final Emo-Rec Point with the widely utilized Face Scale, communicate effectively with facility nursing staff, and facilitate further discussion, we set the range to 1–20. Additionally, this weight was set to filter out emotions other than happiness and closely align with the Face Scale. The Emo-Rec Point formula enabled us to quantify the emotions of the RGT participants and conduct subsequent analyses.

Software Tool Used in the Experiment

We use Python 3.7 to implement the Emo-Rec Application of this research, mainly based on GitHub: https://github.com/oarriaga/face_classification. Opencv-python:4.1.0.25; tensorflow:1.13.1.

3.3 Validation of Recognition Accuracy by Questionnaires

3.3.1 Object and Methods

This thesis aimed to determine whether an AI image recognition system can accurately assess the emotional abilities of older individuals with dementia through a survey completed by fully experienced caregivers.

Each session of the RGT program typically lasted 15 min. The accuracy of the observation results was affected by the shooting angle, motion state, and accurate quantification of the time point. It is not feasible to identify emotions and extract data solely based on the duration of each experiment. Instead, we can utilize the video data collected during the RGT experiment, which can be directly employed in the Emo-Rec Application. We began by selecting appropriate time points, recording the frame time of the current video, and scoring and documenting the current frame image.

Nine time frames were marked in each RGT session video following a random sequence supplied by SPSS 12 Compatible. The frames of each video were categorized using the Emo-Rec Application to obtain the proportion of each emotion expressed by the participants. The Emo-Rec Point was then calculated as shown in Figure 3-3. Further details on the RGT are provided in Section 4.3.1.

It is challenging to accurately capture emotional changes over time, and the speed and direction of change cannot be predicted with certainty [27]. Multiple sampling can yield a relatively accurate representation of the overall emotional tendencies of RGT participants during each session [29]. To achieve this, we marked time frames in the RGT videos with reference to the random sequence provided by SPSS 12 Compatible and classified frames using the Emo-Rec Application to determine the proportion of emotions expressed by participants. Based on the video length, we created a set of random numbers to select specific seconds for sampling. SPSS 12 random number generation technique was used. For instance, if the video material was ten minutes long, nine random numbers between 1 and 600 were generated. The time frames were then extracted using the seconds associated with the random numbers to complete the sampling.

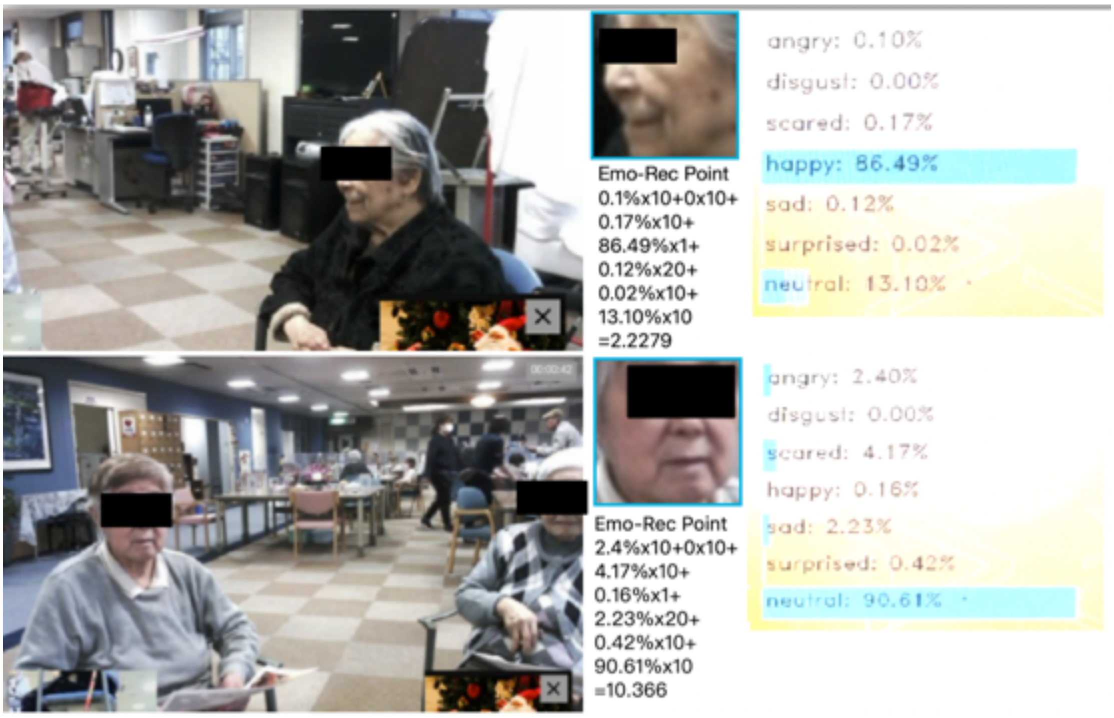
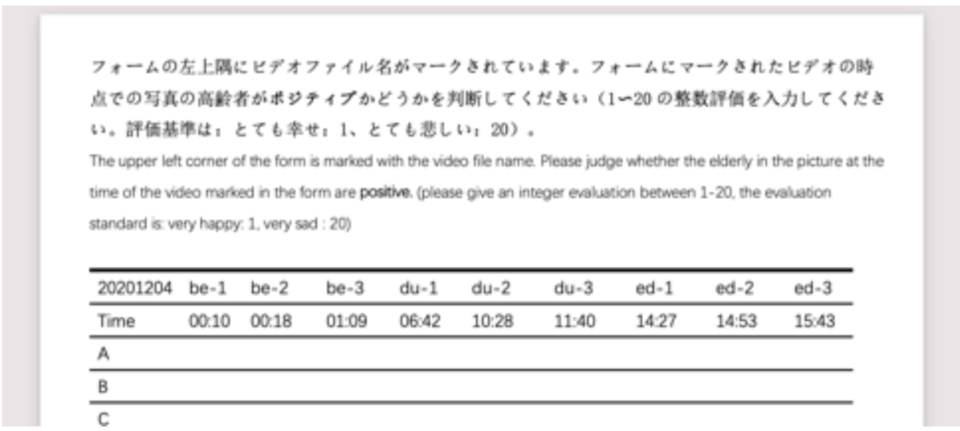


Figure 3-3. Obtain the Emo-Rec Point from the Video Data

To verify the precision of the Emo-Rec Point, we recruited three professional nursing staff and experts specializing in nursing older PWD. They completed questionnaire surveys to evaluate the recorded time points of the videos acquired using the Emo-Rec Application, as shown in Figure 3-4. We selected three participants based on specific criteria requiring extensive experience in relevant clinical or research studies conducted in East Asia. Additionally, two experts had backgrounds as scholars in social welfare, while the third was a member of the nursing staff at an older adult care facility in Japan.



フォームの左上隅にビデオファイル名がマークされています。フォームにマークされたビデオの時点での写真の高齢者がポジティブかどうかを判断してください（1〜20 の整数評価を入力してください。評価基準は：とても幸せ：1、とても悲しい：20）。

The upper left corner of the form is marked with the video file name. Please judge whether the elderly in the picture at the time of the video marked in the form are **positive**. (please give an integer evaluation between 1-20, the evaluation standard is: very happy: 1, very sad : 20)

20201204	be-1	be-2	be-3	du-1	du-2	du-3	ed-1	ed-2	ed-3
Time	00:10	00:18	01:09	06:42	10:28	11:40	14:27	14:53	15:43
A									
B									
C									

Figure 3-4. Questionnaire Surveys for Participants

Subsequently, we summarized the questionnaire results to confirm the coherence of the opinions of the three participants. We compared the mean outcomes with those obtained from the Emo-Rec Application to determine its effectiveness in recognizing emotions among older patients with PWD.

Ethical Considerations

This interventional approach 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 the participants signed a consent form to participate in this study.

3.3.2 Results

The two groups of RGT intervention experiments were recorded over six sessions with eight PWDs participating. Owing to the physical condition of the participants and the influence of the shooting angle, 258 time points were extracted from the six intervention experiment sessions, with 174 lost due to absence and device issues. The three participants who scored in the same session demonstrated varying scoring patterns in Table 3-1. However, a one-way ANOVA did not reveal any statistically significant differences as Table 3-2 shown.

Table 3-1. Analysis of Surveys

	N	M	Std	p	95%CL	
					LL	UL
P1	258	10.3566	3.47710	.21468	9.9303	10.7829
P2	258	10.6822	3.38912	.21100	10.2667	11.0977
P3	258	10.6395	3.37735	.21026	10.2255	11.0536
Total	774	10.5594	3.41346	.12269	10.3186	10.8003

Table 3-2. One-way ANOVA

	Sum of S	df	MS	F	Sig.
Between Groups	16.158	2	8.709	.693	.500
Within Groups	8990.609	771	11.661		

We used a one-way ANOVA to determine whether the scores of the three participants were consistent and found no significant difference (sig. =0.5, p>0.05). Kendall and Spearman tests were also used to verify the consistency of the two evaluation methods (Table 3-3).

Table 3-3. Kendall's and Spearman's Tests

	N=258
Kendall's Test	.722**
Spearman's Test	0.863**

***. Correlation is significant at the 0.01 level (2-tailed).*

We conducted a kappa test on the outcomes of the two groups, and the findings indicated that there was a moderate level of consistency between the results of the two groups (kappa 0.21–0.40) (Table 3-4).

Table 3-4. Kappa Test

	Value (n=258)	Standard Error	Approximate T	Approximate Significance
The Measure of Agreement Kappa	.039*	.014	5.338	.000

Additionally, because a rating scale of 20 may not be appropriate for the kappa test, the intraclass correlation coefficient (ICC) was utilized as a criterion to determine the consistency between the scores given by the nursing staff and the Emo-Rec Application (Table 3-5).

Table 3-5. Result of Intraclass Correlation Coefficient

	ICC	95% CL Lower Bound	95% CL Upper Bound
Single Measures	0.835	0.642	0.909
Average Measures	0.910	0.782	0.953

In summary, consistency of the Emo-Rec Application and the results reflected by the nursing staff can be assumed. It is foreseeable that more efficient AI systems will be used in this field in the future to obtain statistical data with greater reference values.

3.3.3 3.3.3 Discussion

The Emo-Rec Application offers a comprehensive range of reference data for various emotional dimensions, including surprise, fear, and disgust. The professional expertise of the nursing staff and experts allowed us to verify the applicability of the tool to both positive and negative emotions. In addition, this tool provides an opportunity to analyze additional emotions in future research to obtain conclusive results across a wider range of dimensions. Surprise and disgust are more frequently detected during initial interactions with robots but tend to decrease over time. We further discovered that older adults with fewer teeth were more prone to expressing anger, whereas older PWDs with a higher nursing burden (NPI>30) exhibited greater fluctuations in mood than those with lighter care needs (NPI<30).

The emotions observed in the PWDs can be explained by previous research findings. Principles such as person entertainment (PE) can be used to explain the positive emotions experienced by older individuals with dementia who interact with robots [38]. Studies have indicated that non-pharmacological treatment of the BPSD plays a crucial role in improving the mood of individuals, with emotions tending to be positive under long-term therapy [39-40]. These findings can aid in better understanding of the various mood changes experienced by older individuals with dementia undergoing non-pharmacological therapy in future research.

With the current implementation of the Emo-Rec Application, we were able to accurately identify the emotions of most older individuals with disabilities in a short period. Consequently, the Emo-Rec Application faces significant challenges in its operation. Owing to the ongoing COVID-19 pandemic, wearing masks is mandatory in public social settings. Although some samples in the system exhibit partial occlusion or rotation, the challenge lies in making them work, because the masks typically cover half of the face, as depicted in Figure 3-5.

Ordinary Situation:

Anger:	0
Disgust:	0
Scared:	0
Happy:	30
Sad:	0
Surprised:	0
Neutral:	70

Wearing a Mask:

Anger:	?
Disgust:	?
Scared:	?
Happy:	?
Sad:	?
Surprised:	?
Neutral:	?

Figure 3-5. Masked Face in the Interventional Experiment

The Emo-Rec Application for older individuals with disabilities results in higher accuracy through depth-wise separable convolution technology. It enhances recognition accuracy by increasing the number of training samples and can be flexibly matched through different training models [34-35]. Rachael's research indicates that training databases for Asian, European, and American faces may not be universal and may lead to specific errors. To obtain superior outcomes, it is advisable to use emotional databases from culturally diverse regions [41]. Finally, optimizing the data collection frequency based on the intervention method and time can enhance facial recognition efficiency, leading to a more accurate reflection of the intervention experiment results.

3.4 Facial Emotion Recognition Tool as an Interventional Approach

3.4.1 Object and Methods

This section substantiates the usability of a facial emotion recognition tool for RGT by conducting teleoperated robotic experiments. Two experiments were performed to compare and analyze the expression data collected from the human and robotic interactions. This study focused on identifying the differences in emotion recognition between these interactions. The two robots used in the experiments were Telenoid and Pepper.

The participants were required to possess fundamental communication skills and exhibit lower scores on the Mini-Mental State Examination (MMSE). Experiment T took place at an inexpensive nursing home in Tokorozawa, Saitama, involving six PWD aged 87.5±5.76. The PWD had an average MMSE score of 19 (Range 13–27) (Figure 3-6). From July to August 2018, six PWD engaged in six 20-minute RGT sessions. The study conducted a crossover interventional experiment consisting of three teleoperation robot sessions followed by three human intervention sessions, while the other group underwent the same procedure in reverse order [42].

Experiment N, described in Chapter 4, involved eight individuals with dementia residing in a nursing home in Nakameguro, Tokyo, with an average age of 85.75±3.7 (G1) and 89.00±2.37 (G2), and with MMSE scores of 15.25±4.85 (G1) and 17.25±12.21 (G2). The experiment was conducted from December 4, 2020, to January 12, 2021 (Figure 3-7). The data from the two intervention experiments were compared separately for humans and teleoperated robots using the Emo-Rec Application to measure the emotion ratings of the participants.

Figure 3-6. Telenoid and Experiment T

Figure 3-7. Pepper and Experiment N

Ethical Considerations

The interventional approach complied with the requirements of the Waseda Ethical Review Committee (ethical review number 2019-328(N) 2017-173(T)). The experimenters explained the experimental content, possible risks, and protection of personal information to the participants. All the participants signed a consent form to participate in this study.

3.4.2 Results

The cases in which Emo-Rec points were under 10 by different intervention methods in the two experiments are summarized in Table 3-6. The participants exhibited more relaxed emotions when faced with the teleoperated robot. The proportion of face-to-face participants with a total score < 10 points was lower than that of the robot group participants. Finally, Experiment T had a slightly higher number of emotionally pleasant people than did Experiment N.

Table 3-6. The Emo-Rec Point of Interventions (< 10)

Experiment		< 10	% Within	X^2	p
T (n=258)	Human	112	69.1%	0.057^{a}	0.811
	Robot	110	67.9%		
N (n=324)	Human	56	44.4%	0.596^{b}	0.440
	Robot	65	49.2%		
Human in Total		168	58.3%	0.085^{c}	0.770
Robot in Total		175	59.5%		

*0 cells (0.0%) had an expected count of less than 5. The minimum expected count was 51.00^{a}. 59.09^{b}. 118.27^{c}.

The average Emo-Rec points of all the participants in each RGT were aggregated in both experiments, as presented in Table 3-7. We then summarized the results in a trend graph to observe the effects of the two different experimental settings on the participants' emotions. Figure 3-8 illustrates that the trend for Experiment T was relatively flat compared to that of Experiment N. The group means, in general, were predominantly below 10, indicating that the RGT was a calming and pleasant experience and that the moods of the participants exhibited signs of improvement over time.

Table 3-7. The Overall Average Emo-Rec Point of Participants

		G1	G2	G3	G4	G5	G6
Experiment T	Human	8.687	9.129	9.191	9.446	8.078	8.739
	Robot	8.561	10.052	7.298	7.913	7.941	8.539
Experiment N	Human	-	11.091	9.759	10.711	8.335	10.576
	Robot	9.683	10.663	-	7.959	9.026	6.486

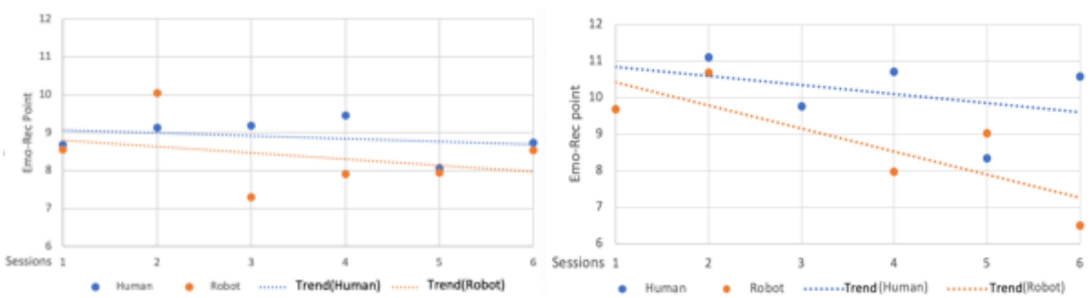


Figure 3-8. Trends of Experiments T & N

3.4.3 Discussion

Based on the data in Table 3-6 for points under 10, there was a significant increase in the number of people experiencing pleasure in the crossover setting of Experiment T. However, the robot group outperformed the human operator group in Experiment N of the controlled trial and overall data analysis. This indicates that interacting with a robot elicits a stronger positive emotional response than interacting with a human operator. This is in accordance with Heerink et al.'s concept of "Perceived Enjoyment" where users experience pleasure when choosing to use a robot [38]. To verify the relationship between user enjoyment (Perceived Enjoyment) and the willingness of older adults to use robots, Heerink et al. conducted a robotic experiment. According to Heerink et al.'s theories, Perceived Enjoyment is a crucial factor in the decision to use robots. This indicates a relatively positive affective trend observed in the robot group during the experiment. Additionally, robots possess hedonic effects, such as new technologies and products, which tend to enhance mood.

Participants in Experiment T underwent a crossover setting in which the impact of alternating stimuli was more pronounced, leading to a significantly more relaxed mood. Experiment N employed a single stimulus in a controlled trial, with both groups receiving human or robotic interventions. Participants in Experiment N scored lower on their overall Emo-Rec points.

Figure 3-7 depicts a relatively flat trend for Experiment T compared to Experiment N. Different interventional settings have diverse outcomes. Overall, the mood of the participants improved during long-term RGT, with greater improvement observed in the robot group than in the experimental group. These findings demonstrate the ability of the tool to analyze different experimental setups.

In a related study, Yamazaki et al. described the therapeutic benefits of the Telenoid robot in facilitating conversations among PWD to improve dementia symptoms [7]. Shibata and Wada considered Face sSale enhancement crucial for reducing BPSD in teleoperated robot studies. This tool simplifies emotional improvement and aids caregivers in adopting cognitive-enhancement strategies [19]. Nevertheless, the performance of facial recognition applications in terms of the attention metric is inadequate. In the future, it is anticipated that eye movement recognition equipment or other sensors will efficiently capture this type of data and offer researchers multidimensional references.

3.5 Facial Emotion Recognition Tool for a Mute Older Adult

3.5.1 Object and Methods

This section aims to describe a nonverbal PWD in intervention experiments with episodes and the trend of the Emo-Rec Point.

During the sessions of interventional Experiment N, we found that participant Mr. I in the control group (human-coordinated RGT) could not speak because of a tracheostomy. Mr. I was able to participate in RGT discussions emotionally and was engaged but could not use linguistic analysis to determine the improvement. Mr. I participated in RGT five times. During the sessions, he actively expressed his words through gestures a few times, primarily through expressions and body language feedback, for group discussions.

RGT aims to enhance BPSD through language-based communication, although the presence of a sample makes it challenging to apply linguistic analysis [43]. Evaluating the impact of participating in

RGT in this unique case using facial emotion recognition may generate insights for future cases and offer novel strategies to on-site staff who encounter individuals with communication difficulties to serve as "social interaction" providers. Subsequently, based on our review, we discovered that the inclusion of a male participant, Mr. I, in daily facility care was rare. We aimed to independently analyze Mr. I's data and investigate the impact of communication therapy on nonverbal PWD. Our study aimed to assess the efficacy of a talking intervention experiment in elderly patients with dementia who were unable to speak, through a multidimensional approach to observation and data analysis.

The Episode of the Mute PWD

During the first RGT session on Dec 4, 2020, Mr. I had a conversation with Ms. T.

"You are still very young; looks like 30 years younger than me."

Mr. I was pleased and responded with a positive gesture. Afterward, Ms. T actively inquired about their living addresses and communicated for approximately three minutes. In addition, he smiled and nodded when the volunteer used a reminiscence card for the RGT.

Owing to the build-up of phlegm in his trachea, Mr. I temporarily withdrew from therapy and received treatment from the facility's nursing staff at the start of the fourth RGT session on December 15, 2020.

Following the interventional study, we conducted a robotic intervention session in the control group. Mr. I struggled to communicate effectively with the robot and maintain prolonged eye contact, prompting him to seek assistance from the researchers on multiple occasions.

Ethical Considerations

This interventional approach 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 the participants signed a consent form to participate in this study.

3.5.2 Questionnaire and Emo-REC Result for a Mute Older Adult with Dementia

According to the MMSE, NPI-NH, and Dementia Behavior Disturbance (DBD) questionnaires, the overall cognitive status of Mr. I remained stable. In terms of MMSE scores, there were no significant differences between Mr. I and the other participants who had undergone therapy. While the NPI results reported a decrease in dementia severity, Mr. I's physical condition increased the workload of the nursing staff. The evidence supporting this perspective stems from the decrease in the level of harm that dementia inflicts, as indicated by the results of the DBD survey presented in Table 3-8.

Table 3-8. Questionnaire results of Mr. I

MMSE		T0	18.00
		T1	18.00
NPI-NH	Severity	T0	6.00
		T1	4.00
	Care Depression	T0	5.00
		T1	16.00
DBD		T0	14.00
		T1	10.00

*T0=before intervention, T1=after intervention

The method using the Emo-Rec Application also yielded a satisfactory improvement. Over the five sessions, the emotion detection results for Mr. I remained relatively steady. The Emo-Rec Application showed a fluctuation range of N=15, mean value of 10.127, and standard deviation of 4.582. According to the calculation method (Emo-Rec Point), lower values correspond to more positive emotion detection outcomes. It can be inferred that the emotional detection data for Mr. I throughout the interventional experiment generally leaned towards positivity as Figure 3-9 shown.

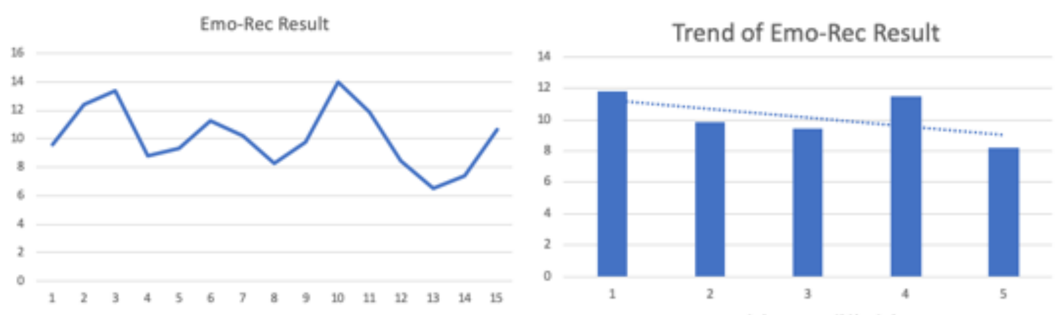


Figure 3-9. Emo-Rec Point results for Mute PWD by Summary/Trend

The results indicated a consistent decline in the Emo-Rec scores in both groups during RGT. Figure 3-10 shows that the mute PWD had an Emo-Rec score ranging from 11.76 8.20. The robot group exhibited a sharper decline than the human group during the RGT, and the mute PWD also displayed a significant reduction in the Emo-Rec score. Both groups demonstrated evident emotional advancement, with a more remarkable trend recorded in the mute PWD and robotic groups.

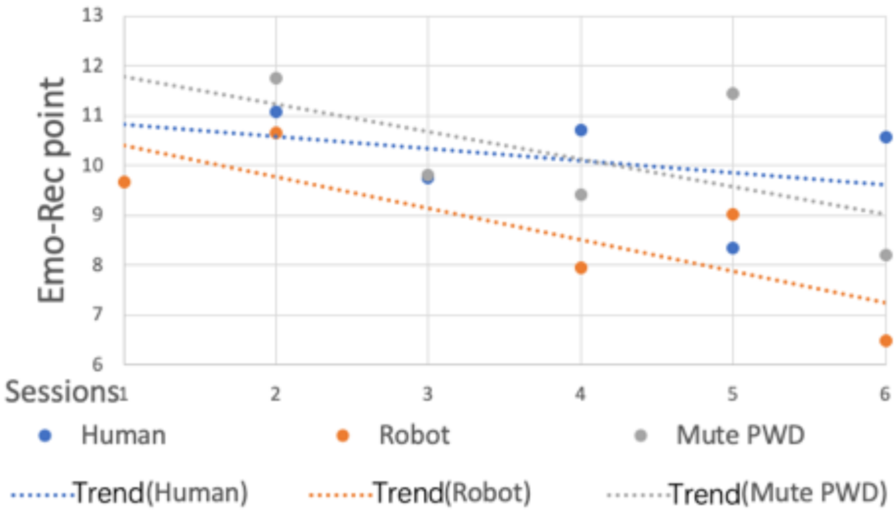


Figure 3-10. Emo-Rec Point Trend of Overall Participants

3.5.3 Discussion

In this study, the primary mode of communication between humans and robots was conversation transmission, as communication through gestures between humans and robots is currently impossible. Even if Mr. I gestures toward the machine, it is challenging to receive the desired feedback promptly. This can cause a certain degree of distress in older people who are unable to speak [44].

Accompanied by future technological developments, improvements can be made to the application of gesture tracking (uSens, Leap Motion, and Oculus NimbleVR) and low-latency network technology for human-computer interaction. Moreover, when combined with a robot control system, gesture recognition, logical control, and action reproduction can be achieved for a more robust outcome. It could offer additional communication possibilities for this type of human-computer interaction technology in future communication between robots and individuals as Figure 3-11 showed [45].

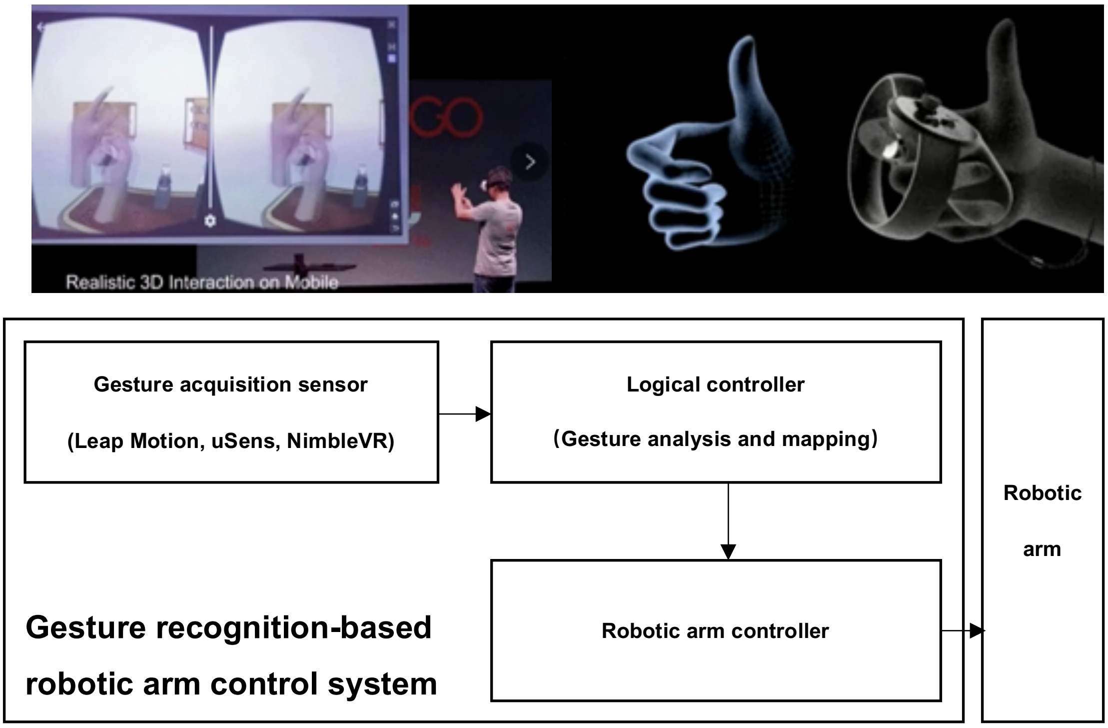


Figure 3-11. Spatial Gesture Recognition and Hand-held Gesture Recognition Equipment

During the RGT implementation in the facility, it was discovered that the communication efficiency of the robot was affected by a limited viewing angle. In addition, communication is affected by factors such as camera resolution and transmission rate.

Nevertheless, through the overall RGT results and observational data, some improvements can be observed in mute individuals with disabilities under current conditions. Notably, the overall trend of the Emo-Rec points decreased from approximately 12 to 9. Objectively, the study's primary finding, alongside other scales, indicates that the RGT intervention brings intuitive improvement to mute PWD in the form of gradual relaxation from tension and mistrust, as well as an increased display of joyful emotions.

3.6 Summary

In this chapter, we endeavor to verify the accuracy and usability of the Emo-Rec Application through a survey of experienced individuals and data from two intervention experiments. The Emo-Rec Application can effectively quantify facial expression data, facilitating the intuition of nursing staff and researchers to evaluate the efficacy of non-pharmacological therapy in enhancing the QoL of older patients with PWD. Additionally, we discovered that nonverbal persons with dementia could benefit from the RGT process during intervention trials.

Compared with previous emotion quantification tools, the Emo-Rec Application enhanced the detection efficiency while maintaining some accuracy.

Our analysis of long-term interventions implies that a facial emotion recognition tool is a more objective way to quantify the emotions of the participants, eliminating the need for alternative instruments. It is simpler and more user-friendly than contact-based electroencephalograms (EEG) and blood tests, and its accuracy can be improved by modifying the feature library of the training datasets. Most importantly, it does not encounter cognitive barriers such as self-reports, questionnaires, or interviews.

In addition, older individuals with disabilities who are unable to communicate verbally may be better suited to emotional assessments using facial recognition technology. This suggests that teleoperated robotics can enhance the emotional state of older individuals with speech impairments. We recommend further improvement of such applications using more precise databases and higher-performing training models to facilitate aid for older PWD in the future.

Chapter 4 Reminiscence Group Therapy for PWD by Teleoperated Humanoid Robot

4.1 Background and Purpose

While conducting the SAR scoping review, we learned that research reports on caring for older adults and dementia related robots have primarily focused on pet-type robots. However, there are relatively few reports on SAHRs. Abdi et al. pointed out that judging how robots can help dementia patients with BPSD more effectively can be evaluated in terms of human imitation, social interaction, and companionship [15,46]. Giraff asked the families of participants to communicate with PWD (n=5) through robots and supporting software to show whether SAR could be used in long-term nursing, which resulted in a satisfactory result. Moyle et al. also pointed out the limitations of using this type of robot in older adult care facilities, which involved a particular learning cost for family members, and there were specific software and hardware compatibility and network stability issues during the experiment. In addition, the study found that evaluation of facial expressions in the study of PWD is an excellent nonverbal expression evaluation method [21].

Building on the above research, we compared and discussed the dimensions of facial expression recognition. We then verified whether deep learning facial emotion recognition applications could correctly recognize the emotions of PWD through the judgment of older adult facility staff and nurses and arrived at a positive conclusion [47].

One method for assessing the impact of a robot intervention experiment is a thorough language evaluation. This approach can objectively showcase the impact of the intervention experiment through quantitative data analysis. A linguistic analysis of the RGT intervention experiment, which employed a Telenoid robot to interact with PWD (n=6), revealed that PWD engaged in more human conversations on average than with the robot. There are no discernible differences in the complexity or richness of the languages used [42].

Pepper is the first social-humanoid robot capable of recognizing faces and basic human emotions. Pepper is optimized for human interaction and can communicate with people through conversations and touch screen. As a comparatively mature robot integrated into homes and businesses, Pepper is highly valued for its future expansion and research. In recent years, social operations related to distance learning and voice calling have become convenient. Engineers continue to develop excellent applications of the Pepper Innovation Challenge to play a role in more work scenarios. During the COVID-19 pandemic, people became increasingly aware of the importance of virtual interactions. Virtual interactions between humans and robots can be applied to more scenarios than the exclusive domains of young people. Dialogue communication through electronic devices may become closer to face-to-face communication, and the boundaries between virtual and reality may become blurred.

Research using Pepper has many similar applications, for example, the use of cognitive stimulation therapy (CST) to support rehabilitation training and cognitive rehabilitation of PWD [48]. Additionally, there were other short-term contacts and good feedback on Pepper's intervention research cases. However, previous evaluations of Pepper's AI dialogue system in senior facilities have been unsatisfactory [42]. The early version developed was too strict for human-machine interaction, and the dialogue content was too simple for long-term communication. However, Pepper's development environment can support the function of remote video calls from the operator; thus, it can function similarly to Giraff, Telenoid, and any other socially assistive robot through software and hardware redevelopment [49].

Evaluating the effects of robot intervention experiments using linguistics is an advanced approach. This method objectively demonstrates the effects of intervention experiments using quantitative data analysis [50]. Linguistic analysis of the RGT intervention experiment using a Telenoid robot for PWD (n=6) indicated that they had a higher average number of human conversations than the robot. However, there were no significant differences in terms of language complexity or content richness. In addition, one of the difficulties in evaluating teleoperated robots is the need for quantitative analysis. According to the literature reviews and previous studies, the number of participants in this type of research ranges from two to three and is still at an early stage [42].

Based on the above research and literature review, the Pepper robot developed by Softbank was used in this intervention experiment [51]. Some developments have enabled remote video communication. In older adult care facilities, the RGT method has been widely used to provide better dementia related BPSD relief services for those in daily care. In Japan, work pressure on nursing staff is high due to labor shortages. From the perspective of labor savings, the RGT method is currently more feasible, reducing the pressure on caregivers and providing more work opportunities for the initial older adult population [52]. When able-bodied older adults become operators, they need to read a simple manual and easily communicate with the PWD according to the theme of RGT. For the broader application of new technologies, the remaining shortcomings of teleoperated robots in replacing the labor force of PWD care need to be clarified. This study used Pepper as an SAHR to implement teleoperated functionality and applied it to PWD in a controlled trial study. The findings are based on observation-based scales, adding linguistics and emotion recognition as objective indicators.

4.2 Methods

4.2.1 Field Setup of the Experiment

We conducted an RGT intervention study using Pepper in an older adult care facility in Nakameguro Ward, Tokyo, Japan. The trial used more wireless connections than wired connections to ensure cleanliness of the facility and avoid the risk of accidental falls by older adults. Because of the viewing angle of the robotic camera, 3-4 participants per RGT would be the ideal configuration. Eight participants from the facility were screened, selected under restrictive conditions and divided into two groups: conventional RGT and robotic RGT, as shown in Figure 4-1. Restrictive conditions included lack of willingness to withdraw from daily care services during the trial, essential communication and cognitive abilities, and willingness to participate in the trial.

We completed the screening and grouping of the participants with assistance from the daily care supervisors at the facility. Eight of the fourteen participants were randomly selected to participate in the study based on the criteria of being daily care users in the facility and having varying degrees of dementia. The supervisor randomly selected eight sticky notes with the participant numbers and placed them in the robotic and traditional communication groups as shown in Figure 3-1. A t-test was used on the questionnaire scale and an independent sample t-test was used in the linguistic analysis to compare the differences in the intervention experiments. Finally, the results of the AI tool are summarized and discussed using trend plots.

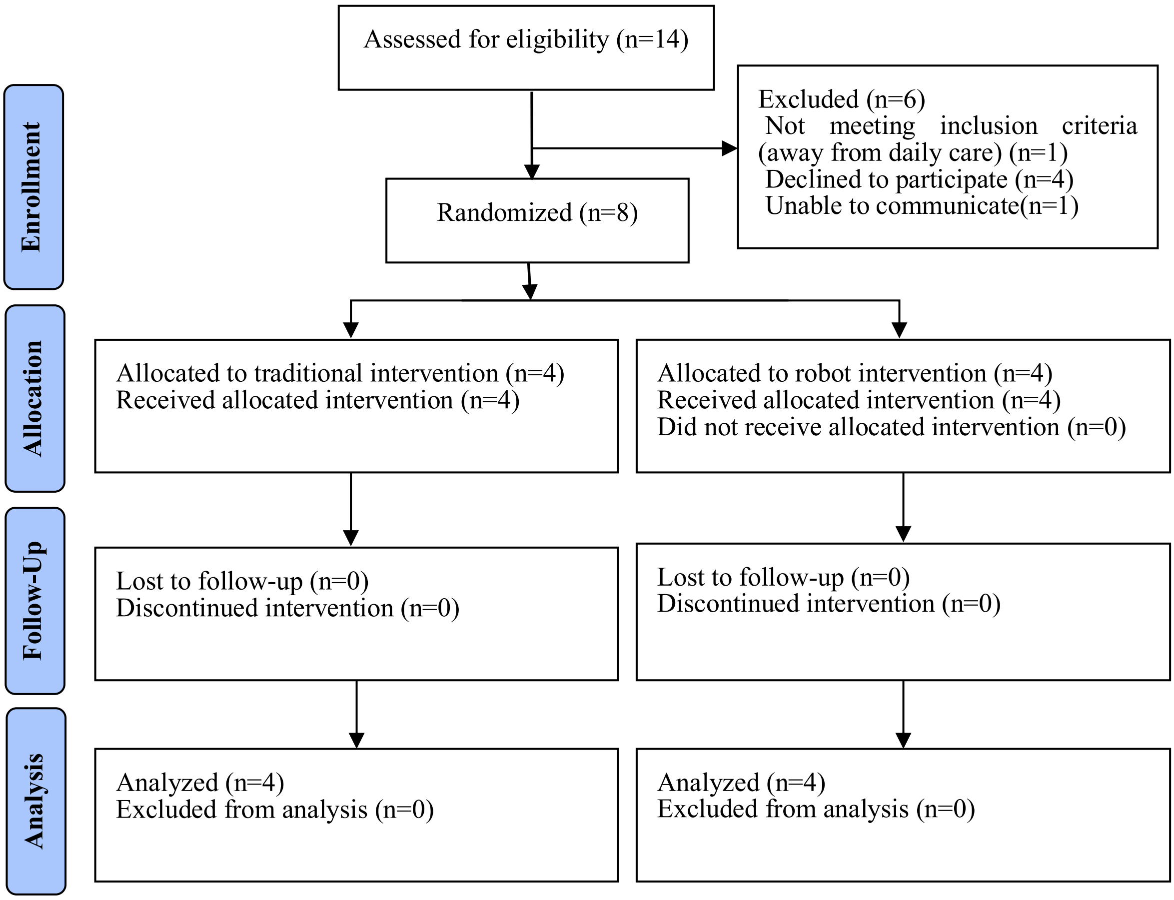


Figure 4-1. The Flow of the Intervention

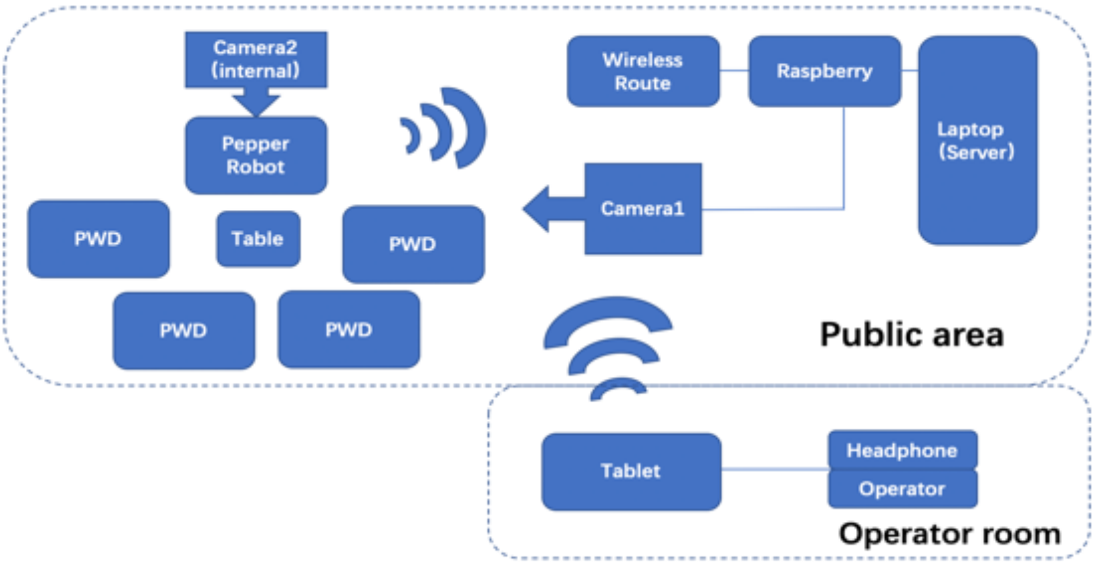


Figure 4-2. Field Setup in the Facility

4.2.2 Interventional Settings

The RGT settings are one-to-many. With the help of the staff at the older adult care facility, we set up a group of four PWD for a two-group intervention experiment (n=8). The mean MMSE score was 18.375 (SD±7.60); the participants were considered to have mild to moderate cognitive impairment (MCI). Among the eight participants, one was generally unable to speak after a tracheotomy, and the effect could not be evaluated using a linguistic analysis. We added the facial emotion recognition dimension to assess the specific differences between the two intervention methods.

We used Raspberry Pi as a remote-control transmission station to help the tablet connect to the loudspeaker and the second camera. We then connected the Raspberry Pi to the server and wireless router. The server provides an IP address to communicate with the operator, so that the Pepper side can communicate. These three parts of the device are connected by wireless networks, as shown in Figure 4-2. Through the tablet, we can display any desired image on Pepper's display screen as Figure 4-3 showed.

The experiment was conducted in a public residential facility for older adults. The participants were divided into two groups according to the facility schedule. One communicated with a human coordinator and the other communicated with the same human coordinator through Pepper. The human coordinator was a 70-year-old local Japanese volunteer who had more in common with the participants than with young people.

From December 4, 2020, to January 12, 2021, all participants conducted six RGT sessions of 20 min each. Each session began with greetings and introductions to help participants learn each other's names. Subsequently, the participants viewed several nursing recreation memory cards with specific numbers, interacted with the contents of the cards, and refreshed the cards for the next session (Figure 4-4). The nursing recreation memory card for RGT is a solution to problems faced by field workers who do not know how to communicate with older adults. The two groups of participants were given the same numbered cards in the same sequence of sessions to control for variables.

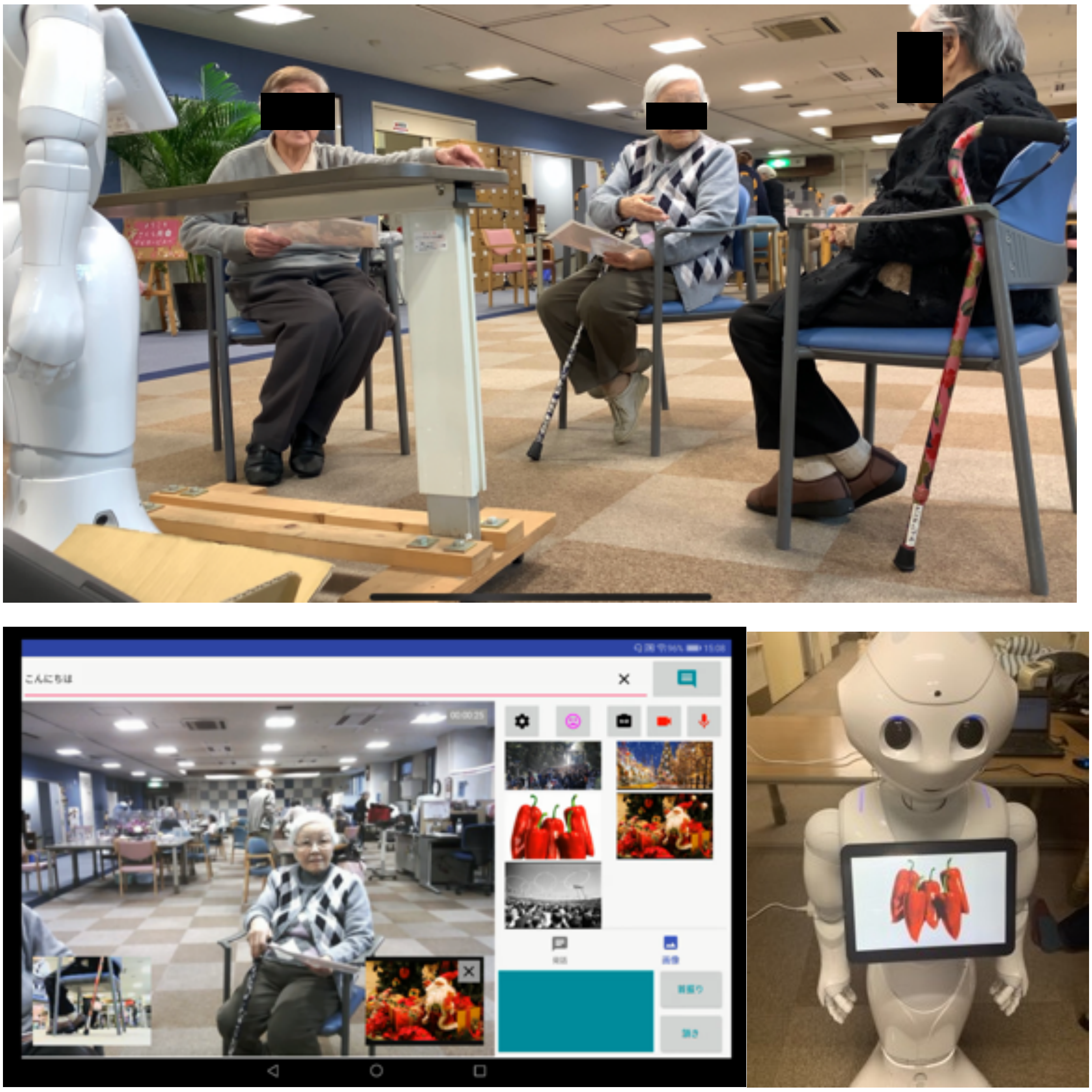


Figure 4-3. Vision in Camera1, Camera2 (tablet) & Picture on Pepper

Figure 4-4. Reminiscence Card for RGT

4.2.3 Participants

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

4.2.4 Data Collection

We administered the MMSE to participants directly. The MMSE is a widely used questionnaire that reflects the cognitive status of PWD [53]. With the help of nursing staff in the nursing home, we completed the Neuropsychiatric Inventory-Nursing Home Version (NPI-NH) and DBD questionnaires before the intervention experiment on December 4, 2020 (T0) and once again after the intervention experiment on January 12, 2021 (T1). The NPI-NH and DBD represent caregiver assessments of PWD, including assessments of physical status, cognitive status, and dementia burden. The MMSE, NPI-NH, and DBD scales are included in the Appendix.

We then recorded all the RGT trials with a voice recorder and video camera, transcribed them into text, and used them for subsequent linguistic analyses. By recognizing the facial expressions, we divided the emotional changes and trends of each RGT participant. Nine video frames were randomly selected for each stage, and the Emo-Rec Application scored nine points on the video. The randomization process used SPSS 12 calculations to determine the number of seconds in the video and to extract the frames.

4.2.5 Data Analysis

For the MMSE, NPI-NH, and DBD results, we used an independent samples t-test to compare the participants' cognitive levels and nursing status improvement before and after the intervention experiment.

We recorded all conversations during the session and divided them into two methods, linguistic analysis and emotion recognition analysis, using IBM SPSS Statistics software (R26.0.0.0 64bit version).

Linguistics Analysis

(1) Utterance and Sentence-Ending Particles

The linguistic analysis includes the statistical analysis of sentence-ending particles ("yo," "ne"). For this purpose, we used the 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 everyday Japanese conversations as well as natural and relaxed conversations [54]. We attempted to identify the possibility of using robots as a substitute for humans in therapy using an independent sample t-test to analyze the differences in tone and wording between PWD in the faces of humans or robots. In addition, we calculated the proportion of the sentence-ending particles 'ne' and 'yo' in the total number of utterances for the two groups. The results are shown in Table 4-3.

(2) Entropy

The other part of the linguistic analysis is the analysis of entropy. The term "entropy" was first proposed in 1948 and refers to the amount of information in a communication whose average value can be measured [55-56]. Measuring the entropy of spoken words can reflect whether a dialogue is persuasive or rich in content. N-gram text, which is used to measure "entropy," has been widely used to evaluate speech and natural language processing tasks [57]. The formula of entropy is shown in Formula 4-1.

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

(4-1)

Entropy (x_i) is the probability of occurrence for item i, log p(x_i) is the amount of information for item i, and the overall formula is the average amount of information for all items. Entropy refers to concepts such as “randomness,” “irregularity,” and “ambiguity.” This reflects the complexity of communication and richness carried in the words [58]. Complexity can be seen as uncertainty when the rules are complex and the expectations (or predictions) of the events that occur are difficult to predict [59].

The word unit entropy analysis method could extract effective semantics of conversations. However, from the data collected during the RGT, we found that the speech content of the PWD participating in the experiment contained more names of people and places, as well as colloquial words. Since the word unit entropy can't fully extract these contents that are not in the standard lexicon, we chose the character unit entropy method in this study. Although the character unit entropy may extract meaningless consecutive characters in some cases, it is able to avoid omissions to capture more meaningful words and phrases than the word unit entropy.

The situation of the participants was reflected by comparing the number of spoken words, sentence-ending particles, and entropy. Therefore, it was necessary to determine whether the participants achieved the required RGT effect in the intervention experiment. We calculated the character unit entropy of the 2-gram, 3-gram, and 4-gram data extracted from the conversation data, as shown in the Results section.

Emotion Recognition Analysis

In this experiment, the Emotion Recognition Application uses a camera to recognize faces and queries a matching database. We believe that the Emo-Rec Application can judge the emotions of patients with dementia using objective data and with higher efficiency.

In the previous section, we compared this type of data with the judgments of the older adult facility staff and nurses to confirm part of the reference value of this tool for practical use. After collecting the average Emo-Rec points for each session, we analyzed the differences between the two intervention methods by trend.

4.2.6 Validity and Reliability

In this study, we discuss the differences between the two intervention methods. It is not sufficiently accurate to judge only the number of participants and facility staff. Therefore, we added two contrasts in linguistics and emotion recognition, and attempted to discuss the differences between robots and humans in RGT from different angles. Counting the utterances in the two intervention experiments can reflect the state of older adults when they face robots or humans. Linguistic analysis objectively reflects whether there is a significant difference in the effect when there is a significant difference in the number of utterances.

Previous standards for evaluating emotions on the Face Scale required participants to express their mood state subjectively. Obtaining data on objectively rated expressions is challenging. Under this consideration, relatively objective results would be accepted if the improvement in emotions were measured using deep-learning AI. We used a training model integrated into the FER2013 dataset (another relatively complete classification database). The classification accuracy is 74%. The ICC between the Emo-Rec Application and the scores of the nurses was 0.835, which can be considered consistent [47].

4.2.7 Ethical Considerations

This study complies 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 the participants signed a consent form to participate in this study.

4.3 Results

There were no statistically significant differences in the scores of each scale before and after the intervention. The results showed no differences between the two groups in terms of level of cognition and caregiver burden as Table 4-1 showed.

Table 4-1. Questionnaire Results of G1/G2

		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

*G1 Traditional RGT group; G2 Robot GRT group, T0 before intervention, T1 after intervention

[a] Independent Samples T-test, MD mean difference T1-T0.

4.3.1 Questionnaire Results

Before and after the experiment, we performed MMSE, NPI, and DBD assessments on the two groups of participants, G1 (mean age=85.75, SD±3.7) and G2 (mean age=89, SD±2.73). The deterioration of the physical condition of several participants during the experiment led to an increase in the NPI and DBD indicators. The remaining participants showed varying degrees of cognitive improvement [58]. We used the mean difference (T1 minus T0) between the two groups for an independent sample t-test analysis to compare the questionnaire scores of the participants under different intervention methods (Table 4-1).

Based on the MMSE, NPI (S and CD), and DBD measures, the degrees of dispersion were approximately equal between the two groups. Statistical analysis revealed no significant differences between the two groups ($p>0.1$). This means that there were no significant differences in cognition, caregiver burden, or level of injury between the two groups of participants.

4.3.2 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 particles "yo" and "ne" at the end of each sentence. We also used a t-test to analyze the results of the two groups, as shown in Table 4-1.

We found that the G1 group that communicated directly with the human coordinator in terms of utterances (mean=1046.25, SD=876.5) was significantly higher than the G2 group that teleoperated with the Pepper robot (mean=412.65, SD=525.8, Cohen's d=0.877, r=0.401).

For the sentence-ending particle "ne," there was no significant difference between the two invention methods ($t(6)= 1.663$, $p>0.1$). However, the sentence-ending particle "yo" was significantly different for the two groups ($t(6)= 2.313$, $p<0.05$, Cohen's d=0.837, r=0.386).

In addition, we calculated the proportion of the sentence-ending particles 'ne' and 'yo' in the total number of utterances for the two groups. The results are shown in Table 4-3.

Entropy of Linguistic

We compared the character unit entropy of the sessions during an experiment lasting more than one month. The differences in the calculation results of 2-gram, 3-gram, and 4-gram were extracted(Table 4-

2), and the results of the t-test detected the contrast of the results (Table 4-3).

Table 4-2. The character unit entropy of G1/G2

Session	2-grams		3-grams		4-grams	
	G1	G2	G1	G2	G1	G2
1	3.184	3.212	3.258	3.306	3.253	3.281
2	3.177	3.276	3.257	3.299	3.171	3.263
3	3.226	3.155	3.275	3.228	3.271	3.307
4	3.196	3.214	3.224	3.295	3.299	3.276
5	3.227	3.227	3.288	3.305	3.287	3.286
6	3.243	3.243	3.251	3.245	3.185	3.278

*G1 Traditional RGT group; G2 Robot GRT group

Grams The frequency of a phrase composed of consecutive Japanese pronunciations.

Table 4-3. Linguistics Results of G1/G2 (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*
Ne [a]		42.58±36.99	21.765±30.31	1.663	.108
Yo [a]		14.75±11.89	6.411±7.56	2.313	.029*
Ne/Utterance		0.042±0.019	0.0546±0.02	2.886	.016*
Yo/Utterance		0.0144±0.005	0.0171±0.006	0.728	.483
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

*G1 Traditional RGT group; G2 Robot GRT group; [a] Independent Samples T-test. *p<0.05

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

4.3.3 Emotion Recognition Results

From the first to the last session, the trend of changes in the emotional scores of the two groups of participants were relatively similar as Figure 4-5. From the Emo-Rec score formula, it can be concluded that the lower the score, the happier the participant. In addition, the emotion recognition results for the robot group showed a positive trend.

Through the facial expression data of the two groups of participants using different intervention methods, we found that the differences between the two methods in various dimensions did not change, regardless of whether they were performed by robots or directly by humans. This means that using robots instead of human labor under the premise of RGT can also satisfy participants emotionally as Figure 4-6 shown.

This experiment illustrates the change in the average score of the participants at the beginning was 10.387, which then decreased, ending at 8.531. This indicated that the emotions of the participants improved more positively during the RGT process.

.

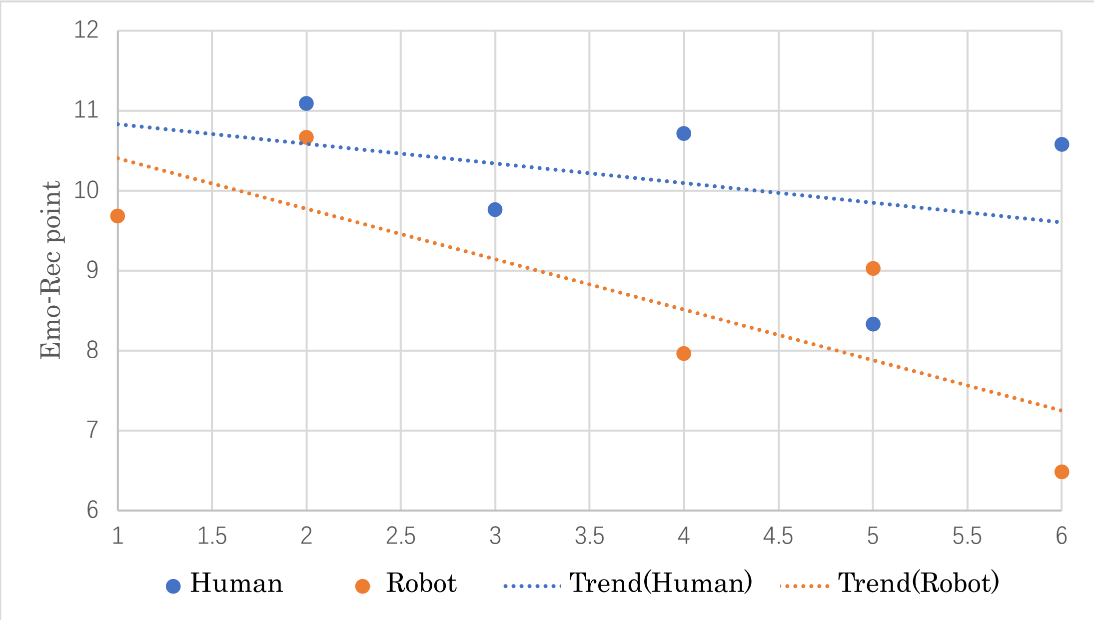


Figure 4-5. The Trend of Emo-Rec Point during RGT

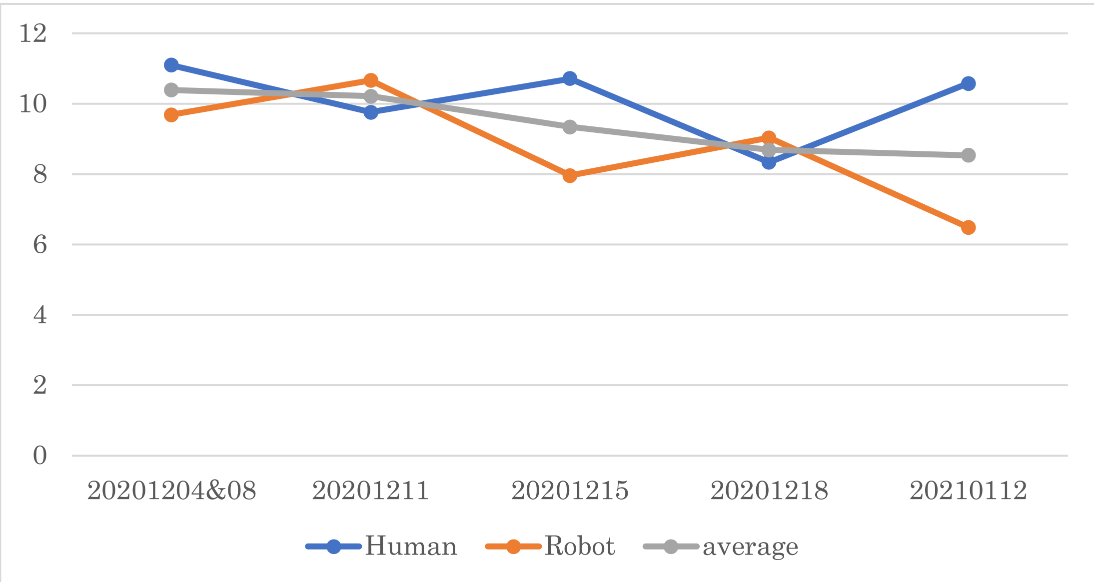


Figure 4-6. Emo-Rec Point during RGT

4.4 Discussion

Treating or alleviating BPSD in PWD generally consumes considerable human and material resources, causing work pressure for caregivers and tremendous inconvenience to their personal lives. If there is a shortage of nursing staff in the future, the use of robots to gradually reduce the workforce will become an important issue. With the help of robot studies in nursing facilities, we designed an intervention experiment in older adult facilities using Pepper robots to determine the remaining shortcomings of teleoperated robots in replacing labor in dealing with BPSD.

We conducted an intervention experiment in a nursing home in Tokyo for more than one month and then compared humanoid robots and human-coordinated RGT to help alleviate the symptoms of BPSD in PWD. The MMSE scale has been widely used in the study of dementia improvement and is a more intuitive and convenient way to reflect the cognitive progress of the participants [60]. The use of the NPI-NH and DBD scales reflected, from the perspective of the facility staff, whether the two intervention experiments reduced the nursing workload at the same level. Meanwhile, the two intervention methods showed no differences in the scale results.

As the intervention experiment was conducted for more than one month, the data reflected by the scales cannot provide more evidence; therefore, we added the analysis methods of linguistics and facial recognition to provide more verification.

4.4.1 Linguistics Analysis

The result was consistent with our prediction: the robot group uttered significantly less than the human group. The teleoperation process reduces the speed of human conversation owing to hardware and network limitations; we also found the same result in another teleoperated robot study [42]. A newly developed experimental system affected by the network quality should be considered in the results. The significant elements of the network Quality of Service (QoS) defined by the Transmission Control Protocol/Internet Protocol (TCP/IP) protocol, packet loss, bit rate, throughput, transmission delay, availability, and jitter, would affect the experimental results to some extent. Under the influence of delay, the efficiency of the utterance between the participants and operator will not share the same place in face-to-face therapy. Voice over IP (VoIP) technology is based on the TCP/IP protocol and has spread to all aspects of life; however, the use of 802.11 wireless protocols still limits the response speed of the signal. In future applications, 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 the QoS.

In Japanese, the sentence-ending particle "ne" is used to seek the consent of others and to empathize with others [61-62]. Although there is no significant difference in the counts of 'ne' between the two groups (t(6)= 1.663, p>0.1), there is a statistically significant difference in the frequency of using 'ne' in the total number of utterances (t(6)= 2.886, p(0.016)<0.05). The common ground between face-to-face and human-machine interactions may be different. Common ground is the knowledge, beliefs, and assumptions that the participants believe they share regarding the activity. It is the basis for all joint actions and is therefore essential for creating the meaning of the speaker and the 'understanding of the receiver [64]. When the dialogue partner is a robot, participants may think that the robot has less in common with humans. Therefore, they used "ne" less often.

However, despite the difference in the number of utterances (t(6)=2.313, p(0.029)<0.05), there is no statistical difference in the frequency of using 'yo' in the total number of utterances between the two groups (t(6)=0.728, p(0.483)>0.05). By mainly using the sentence-ending particle "yo," a speaker reminds the other party to pay attention to things they do not understand or to share information with the other party [63-64]. These results suggest that both groups used the same soft tone, resulting in a relatively similar communicative atmosphere.

The comparison showed that in the character unit entropy result, there was no noticeable difference in the complexity and richness of the content between the two intervention methods. The analysis of character unit entropy was conducted to avoid participants from speaking meaningless content in both interventions due to cognitive disorders that can affect the effectiveness of the therapy, such as serial repetition due to forgetting and no-feedback communication with only intonational words.

The result of character unit entropy (2-gram t=-0.622 p=0.548>0.05; 3-gram t=-1.278 p=0.230>0.05; 4-gram t=-1.648 p=0.130>0.05)implies that the PWD produced the same level of intelligence quantity when speaking to robot or person during the RGT process.

4.4.2 Emotion Recognition Analysis

Although it is difficult for PWD to express their facial expressions, the verification of emotions requires subjective results. SAR intervention tests usually use questionnaires, such as face scales and satisfaction surveys, to demonstrate the feasibility and usability of teleoperated robots [7].

The trend graphs derived from this tool showed that the robotic intervention improved the mood of the participants more significantly. This result is consistent with Heerink's theory that robots make users feel comfortable during the initial stage of a PE intervention. With time, people's confidence in the use of robots also brings a better mood to users [38]. The significance of using an emotion recognition tool is mainly in solving the authorization restrictions and database problems when using the previous software. Previous face recognition software required complicated authorization and usage restrictions, which inevitably caused researchers to extend the experimental period. There are also cases of wasted experimental data due to the limitations 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, owing to the limitation of database matching, intervention research often selects experiments of different races in different regions, which requires the support of local databases to ensure recognition accuracy [41].

This method can improve accuracy by searching for a suitable database, thus providing more valuable research results. To provide better verification ideas in the current situation, robots used to assist PWDs are constantly and rapidly being updated by finding better and more efficient facial recognition methods.

.

4.5 Summary

The linguistic results for the teleoperated humanoid robot group were inferior. In contrast, the PWD of the robot group for mood improved with a better trend. Conversational communication between humans and robots is less smooth or natural than face-to-face communication. Improving this aspect could be an effective measure for robotic applications in cognitive care.

In the process of conversation and feedback, we found that many participants were more active in discussing their shared experiences with their youths. During the experiment, the operator failed to balance the time allocated to participants.

Regarding the selection of the operator for the robotic intervention experiment, we found that older adults who could work were more suitable for this role. This is because large differences in age often lead to difficulties understanding and communicating [65]. Older adults with dementia recognize familiar things better but have a slower acceptance process for strangers and new things.

However, older adult caregivers do not have the advantage of using robots or new technological products compared to the younger generation. With modern networking, caregivers can also remotely control robots. Through network and mobile communication technologies, operators can conveniently access therapy for the disabled at any time. This changes the form of work and provides opportunities for older adults to work.

As a solution to this labor shortage, the teleoperated robotic RGT method is currently feasible, reducing the pressure on caregivers and providing more work opportunities for older adults. They can become operators and need to read the instruction manual to easily communicate with the PWD according to the topic of conversation.

In current Japanese society, the reasons for the slow increase in older adult employment are mainly the following three points:

1. Older people's work experience is valuable, but few jobs can utilize this experience, and older people often need to learn new work skills, which makes reemployment of older people even more difficult.

2. The overall efficiency of the older adult workers is lower than that of the young, and the wages are relatively unsatisfactory.

3. Pension consumption is a more sensible choice for most older adults, and the resulting low willingness to work is also an important reason.

From the perspective of employers and the social environment, mandatory retirement systems, payroll management, and compensation systems are objective factors that make it difficult for older people to participate in enterprises. In addition, there are problems with health risks to older adults and social discrimination [52].

For these reasons, it is difficult to build an employment-friendly social system for older adults. However, replenishing the young labor force, which has a low birth rate and aging population, is a top priority in Japanese society. With improvements in medical conditions, the number of healthy older workers is also increasing, and it is important to create a more employment-friendly social system for all age groups.

Owing to the use of electronic devices, it is necessary to check a simple instruction manual before starting work to properly use the entire set of devices. A 70-year-old volunteer quickly learned and started working; no operational problems occurred during the interventional experiment. This study demonstrates that the use of teleoperated RGTs can create more employment opportunities for older workers. It was also hoped that after the COVID-19 epidemic, mental health care work would be made available anytime and anywhere [66].

In conclusion, further improvements and detailed research are required to identify and solve the problems that current robots cannot perform in dementia care. We should conduct more studies with more methods of using robots to alleviate the labor shortage of caregivers and identify tasks that robots cannot perform and solve.

Summarizing the intervention experiments, we identified several limitations. First, the sample size was small, and it was impossible to draw substantial conclusions by comparing the two groups. We look forward to solving the sample size problem by deepening this research in the future. Owing to equipment and camera angle limitations, some data could not be fully generated, resulting in a lack of facial recognition samples. More convincing research can be conducted using more sophisticated and mature equipment in the future.

Chapter 5 Conclusion

5.1 Summary

This thesis is divided into three sections following the introduction. The first part is a scoping review that summarizes and analyzes existing research related to SAHR used for dementia. The SAR evaluation framework was used to better locate the role of SAHR in the BPSD mitigation. We are currently identifying appropriate research methods for SAHR to assist PWD. From the scoping review, we learned that five roles could be used to design subsequent robotic experiments: affective therapy, cognitive training, social facilitation, companionship, and physiological therapy. In future research, the use of mature mass-produced robots will effectively address this situation. Improvements in robotics technology can also better meet the needs of the field, such as human imitation, social interaction, and companionship.

After the scoping review study, we decided to use the Pepper robot among existing mass-produced robots for subsequent intervention experiments. However, scholars have been excited about the concept of automatic facial emotion recognition. Using Python's deep learning AI and open-source programming tools, we realized an application that can automatically recognize the emotions of PWD and called it the Emo-Rec Application.

The second part involved the divergent verification of the AI tool (Emo-Rec Application) used in the experiment and reflection on the problems found in the field experiment. To verify the usability of the tool, we identified one nurse and two experts with experience in caring for older PWD. They were asked to rate the expressions of the PWD participants in the intervention experiment and compare them with the results of the Emo-Rec Application. An ANOVA test (sig. =0.50) was used to determine whether the judgments of the caregivers deviated significantly. Kendall's test (0.722), Spearman's test (0.863), Kappa test (0.39), and ICC (0.835) were used to verify whether the emotion recognition system was consistent with the judgments of the nurses and experts.

The results of both evaluations were consistent. This implies that this tool can provide an efficient and accurate facial emotion recognition for PWD and provides a theoretical basis for further use.

The AI tool used in this study was also used in other teleoperated robot experiments (an older adult robot dialogue experiment using Telenoid) to verify whether it was consistent with the results of previous studies. We also found that mute PWD could be emotionally enhanced during conversational RGT.

In the third part, we designed an interventional experiment that lasted more than one month to perform RGT therapy on PWD through human-coordinated and humanoid robot teleoperation. A total of eight participants in the experiment were divided into two groups, G1 (mean age=85.75, SD±3.7) for face-to-face RGT and G2 (mean age=89, SD±2.73) for robot-teleoperated RGT.

We then compared the two intervention methods in terms of emotion recognition and linguistic analyses. The score of G1, which communicated directly with the human coordinator in the utterance (1046.25, SD=876.5), was significantly higher than that of G2, which teleoperated with a Pepper robot (412.65, SD=525.8). For the sentence-ending particle "yo," there was no significant difference between the two invention methods ($p>0.1$). However, the sentence-ending particle "ne" was significantly different for the two groups ($p<0.05$). The reason for this was that there might be a difference in the common ground between face-to-face and human-computer interactions.

The entropy analysis results for 2, 3, and 4 grams found no significant difference between the two intervention methods ($p>0.1$). In the t-test results for 4 grams, equal variance was not observed, but there was no significant difference ($p>0.1$). This implies that there was no significant difference in the complexity of communication and richness of content during conversations.

Mood changes improved intuitively with both intervention methods (average: 10.387–8.531). Positive results were observed. This study shows that the use of robots instead of nurses is feasible in terms of nursing care for PWD with BPSD. In addition, this service proposes a new form with high feasibility for employing an older adult workforce.

In summary, we have made it possible to use a facial emotion recognition application based on OpenCV for our experiments. We have overcome various problems in the facilities for dementia elderly people and for general elderly people and developed a RGT method using a remote-controlled robot. We conducted an intervention experiment using this method, collected data, and analyzed it using statistical methods and linguistic methods. To alleviate or and solve the labor shortage problem the nursing industry in Japan is facing now by utilizing the robotics, IT and artificial intelligence technology, this study conducted a multidisciplinary study and experimental exploration in order to investigate the feasibility of our proposed palliative program for people with dementia. Most of the previous studies have been conducted under the laboratory conditions for testing and application, but robot dialogue experiments applied to actual scenarios were rare. In that sense, it is highly expected that the experience of field experiments in the facilities for dementia obtained in this study can provide helpful insights and references for the future research works and practical applications.

The Reminiscence Group Therapy for People with Dementia used in this study is one of the non-pharmacological treatments and is expected to slow the progression of dementia. This cross-disciplinary research has great significance in applying IT and AI technologies, such as humanoid robots and emotion recognition, to the field of health and welfare, and in identifying and finding issues and problems for the social implementation of advanced technologies and research results. The results of this study are expected to lay the foundation for new dementia treatments using intelligent humanoid robots, expand treatment options, and contribute to enhancing the wellbeing of the elderly.

5.2 Future Work

This thesis presents an RGT intervention experiment for PWD using humanoid robots. Attempts have been made to adapt humanoid robot intervention experiments for PWD by summarizing previous SAR evaluation frameworks. We validated the accuracy of facial recognition tools applied to PWD, applied them in several intervention experiments, and determined their applicability to mute PWD. We also attempted to establish evaluation criteria for the use of facial expression recognition tools to perform PWD during RGT. An interventional experiment was conducted in an older adult facility to verify the possibility that humanoid robots can gradually replace human labor in the RGT process. We also look forward to the reforms that the development of AI technology, large language models (LLM), and robotics technology will bring to the field.

When using humanoid robots to alleviate BPSD in PWD, more participants and long-term intervention experiments are needed to obtain more accurate and convincing results. In the future, with the widespread use of humanoid robots, we will obtain more complete data and objective evaluations. We expect that RCTs on humanoid robots will significantly advance the development in this field.

Some companies, such as Boston Dynamics Atlas and Tesla Bots, have invested in humanoid robots, as shown in Figure 5-1. New types of robots will be much more advanced than the Pepper robots in terms of processing speed, sensors, and networking components. It is foreseeable that the experience provided to older adults will be considerable in the advanced future of the robotics industry.

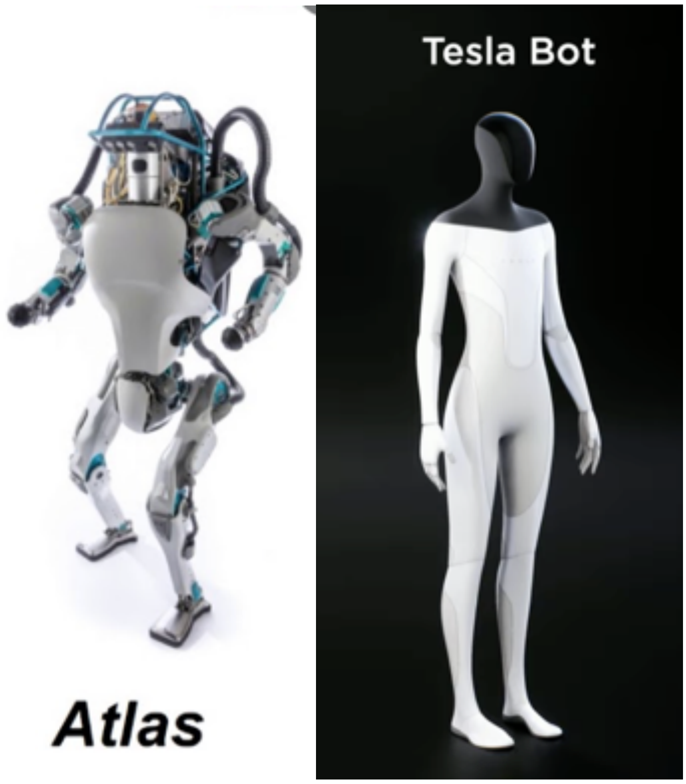


Figure 5-1. Boston Dynamics Atlas and Tesla bot

In addition, the development of wireless technology will directly affect the applications of humanoid robots. The Wifi6/5G local or wide-area wireless technology requires voice or video interactions with lower latency and more explicit images. Humanoid robots that can directly display remote operator actions without delay can add more body language to increase communication complexity [44].

When this study was conducted, owing to the limitations of the immature human dialogue model, we chose the teleoperated form to conduct a BPSD moderation robot dialogue experiment on PWD. In particular, regarding the HCI dialogue program carried out by the humanoid robot, some problems still exist.

1. The response speed could be faster, but the limitation of speech processing performance makes recognition and response speeds unbearable. Although the test personnel showed patience at the beginning and could wait for their responses, they could not maintain their attention during long-term exposure experiments.

2. Thus, the recognition efficiency can be higher. This problem becomes unbearable when most of the subjects in this experiment are older PWD. It has to be repeated many times, and the speech speed has to be slowed down to be recognized.

3. Unable to engage in long-term continuous dialogue. The robot could only respond to a single command, such as what the weather is today but could not continuously have a long conversation about

unknown weather".

The existence of the above problems are the main reasons why we chose to use teleoperated solutions for RGT therapy; however, with the development of the AI technology language model, the above problems can be solved to some extent. ChatGPT, which was developed by OpenAI, was introduced in daily life as a large-scale language model of GPT-3 at the beginning of 2023. Its ability to handle long, continuous conversations raises the prospect of robots with true conversational AI capabilities. However, only a few studies have shown the potential of ChatGPT in medicine, and it remains to be determined whether it can work reliably in various practical application scenarios that require long sessions [67].

Currently, there is a particular gap between ChatGPT and Artificial General Intelligence (AGI). The good news is that now that ChatGPT has attracted widespread attention, large companies are actively directing resources toward this type of technology. We will soon be able to engage in conversations with humanoid robots that communicate smoothly with humans.

Abbreviations in This Thesis

PWD: People With Dementia;

BPSD: Behavioral and Psychological Symptoms of Dementia;

RGT: Reminiscence Group Therapy;

MMSE: Mini-Mental State Examination;

NPI-NH: Neuropsychiatric Inventory-Nursing Home;

DBD: Dementia Behavior Disturbance;

MCI: Mild Cognitive Impairment;

CIH: Cognitively Intact Healthy;

CST: Cognitive Stimulation Therapy;

EMG: Electromyography;

EMFACS: Emotional Facial Action Coding System;

QoL: Quality of Life;

QoS: Quality of Service;

VoIP: Voice over IP technology;

TCP/IP: Transmission Control Protocol/ Internet Protocol;

HCI: Human-Computer Interaction;

AGI: Artificial General Intelligence.

Acknowledgments

First, special heartfelt thanks are given to my respected supervisor Professor Kase Hiroko, whose patient instructions and constructive suggestions were greatly beneficial to me.

Second, I would like to thank Professor Jin Qun, Professor Kikuchi Hideaki, Professor Ogihara Atsushi, Professor Nishimura Shoji, and Dr. Yamazaki skove Ryuji, who taught me and provided generous support during these years. In addition, The analysis from linguistic science was possible because of Professor Kikuchi Hideaki's teaching, that the special thanks belong to him.

I am also grateful to Ms. Yang Jie for helping and conducting a part of the linguistic analysis, and Mr. Ishio Akira for the tele-operation of robots, and to Ms. Nakajima Akiko, Professor Zhao QingBo, Professor Kim YoungHee for their help in the field.

Finally, I would like to express my heartfelt gratitude to my family members who provided me with great help and support. Upon completion of this thesis, I am grateful to those who offered encouragement and support during my studies.

.

List of Published Papers Related to This Thesis

[1] Xiangyu Liu. Social Assistant Humanoid Robots for Dementia of the Elderly: a Scoping Review, Scientific and Social Research. 2020, 2(3)95-103. DOI:10.36922/ssr.v2i3.1001.

[2] Xiangyu Liu. Validation research on the application of Depthwise Separable Convolutional AI facial expression recognition in non-pharmacological treatment of BPSD Journal of Clinical and Nursing Research. 2021, 5(4):31-37. DOI:10.36922/ssr.v2i3.1001.

[3] 劉 相禹：ロボットメディアを活用した認知症高齢者に対する回想法の効果．第 24 回日本在宅ケア学会学術集会 P7B-5 抄録集.2019, p.201.

[4] XiangYu Liu, Ryuji Yamazaki, Jie Yang, and Hiroko Kase. Humanoid robot teleoperation reminiscence group therapy for older adults with dementia: A controlled trial study. Gerontechnology. 2022. 21(1):1-12. DOI:10.4017/gt.2023.21.1.794.02.

[5] X.Y. Liu, R. Yamazaki, H. Kase. How can we measure the effect of reminiscence group therapy in mute individuals with dementia? Trial using facial emotion recognition method. Gerontechnology. 2023. 21(s):671-671. DOI:10.4017/gt.2022.21.s.671.pp4.

[6] XiangYu Liu, Ryuji Yamazaki, Hiroko Kase. Facial Emotion Recognition Tool for Teleoperated Robot Reminiscence Group Therapy for People with Dementia. Gerontechnology. 2023. 22(1):1-6. DOI:10.4017/gt.2023.22.1.827.07.

References

[1] Japan MHLW. Reiwa 2nd Year White Paper on Health, Labour, and Welfare. https://www.mhlw.go.jp/wp/hakusyo/kousei/19-2/

[2] Japan MHLW. Heisei 29 overview of vital statistics. https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei17/index.html

[3] Japan MHLW. Annual Health, Labour and Welfare Report 2022. https://www.mhlw.go.jp/english/wp/wp-hw2022/

[4] Japan Cabinet Office. Annual Report on the Japanese Economy and Public Finance 2018. https://www5.cao.go.jp/keizai3/whitepaper-e/archives/wp2018-e.html

[5] Fuji Economics. Robot Business Corporate Strategy Overview 2021. https://www.fuji-keizai.co.jp/report/detail.html?code=162011842&la=ja

[6] Francois Michaud, et al. Telepresence Robot for Home Care Assistance. Multidisciplinary Collaboration for Socially Assistive Robotics the 2007 AAAI Spring Symposium, 2007, SS07(07):26-28.

[7] Takanori Shibata and Kazuyoshi Wada. Introduction of Field Test on Robot Therapy by Seal Robot, PARO. Journal of the Robotics Society of Japan, 2011, 29 (3):246-249. DOI:10.7210/jrsj.29.246

[8] Lihui Pu, et al. The Effectiveness of Social Robots for Older Adults: A Systematic Review and Meta-Analysis of Randomized Controlled Studies. The Gerontologist, 2019, 59(1):e37–e51. DOI: 10.1093/geront/gny046

[9] Lee, O. E. K, et al. Socially Assistive Humanoid Robots: Effects on Depression and Health-Related Quality of Life among Low-Income, Socially Isolated Older Adults in South Korea. Journal of Applied Gerontology, 2023, 42(3):367–375. DOI:10.1177/07334648221138283

[10] Zuschnegg Julia, et al. Humanoid socially assistive robots in dementia care: a qualitative study about expectations of caregivers and dementia trainers. Aging and Mental Health, 2021, 26(6):1270-1280. DOI:10.1080/13607863.2021.1913476.

[11] Kase H, Analysis and Evaluation of Dementia Care Management Focusing on Behavioral and Psychological Symptoms (BPSD). Waseda University Graduate School of Human Sciences. 2013. (in Japanese)

[12] Finkel S. Introduction to behavioural and psychological symptoms of dementia. International journal of geriatric psychiatry, 2000, 15(1):S2–4. DOI:10.1002/(sici)1099-1166(200004)15:1+3.0.co;2-3.

[13] Izumi Kondo, Shota Suzumura and Aiko Ohsawa. Rehabilitation Medicine for Dementia：Approach According to Its Stage and Introduction of Intelligent Technology, The Japanese Journal of Rehabilitation Medicine, 2018, 55(9):767-772, DOI:10.2490/jjrmc.55.767,

[14] Chia-Jung Hsieh, et al. Reminiscence Group Therapy on Depression and Apathy in Nursing Home Residents With Mild-to-moderate Dementia. Journal of Experimental and Clinical Medicine, 2010, 2(2):72-78. DOI:10.1016/S1878-3317(10)60012-5.

[15] Abdi J, et al. Scoping review on the use of socially assistive robot technology in elderly care. BMJ open, 2018, 8(2):e018815. DOI:10.1136/bmjopen-2017-018815

[16] Cohen-Mansfield J., et al. Can persons with dementia be engaged with stimuli?. The American journal of geriatric psychiatry. official journal of the American Association for Geriatric Psychiatry, 2010, 18(4):351–362. DOI:10.1097/jgp.0b013e3181c531fd.

[17] Moyle W., et al. Use of a Robotic Seal as a Therapeutic Tool to Improve Dementia Symptoms: A Cluster-Randomized Controlled Trial. Journal of the American Medical Directors Association, 2017, 18(9):766–773. DOI:10.1016/j.jamda.2017.03.018.

[18] Derek McColl, et al. Brian 2.1: A socially assistive robot for the elderly and cognitively impaired. IEEE Robotics and Automation Magazine, 2017, 20(1):74-83. DOI: 10.1109/MRA.2012.2229939.

[19] R. Yamazaki, et al. Teleoperated android as an embodied communication medium: A case study with demented elderlies in a care facility. 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. 2012:1066-1071. DOI: 10.1109/ROMAN.2012.6343890.

[20] Momotaz Begum, et al. Collaboration of an assistive robot and older adults with dementia. Gerontechnology, 2015, 13(4):405-419. DOI:10.4017/gt.2015.13.4.005.00.

[21] Wendy Moyle, et al. Connecting the person with dementia and family: a feasibility study of a telepresence robot. BMC Geriatr, 2014, 14(7). DOI:10.1186/1471-2318-14-7.

[22] Granata C., et al. Robot services for elderly with cognitive impairment: testing usability of graphical user interfaces. Technology and health care: official journal of the European Society for Engineering and Medicine, 2013, 21(3):217–231. DOI:10.3233/THC-130718.

[23] Ya-Huei Wu, et al. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the Living Lab setting. Clin Interv Aging. 2014 9:801-811. DOI: 10.2147/CIA.S56435.

[24] R. Q. Stafford, et al. Improved robot attitudes and emotions at a retirement home after meeting a robot, 19th International Symposium in Robot and Human Interactive Communication,

Viareggio, Italy, 2010, (2010):82-87. DOI: 10.1109/ROMAN.2010.5598679.

[25] Strandbech J D. Tema 2: Ethel and her Telenoid: Toward using humanoids to alleviate symptoms of dementia. Tidsskriftet Læring Og Medier (LOM), 2015, 8(14). DOI:10.7146/lom.v8i14.21990

[26] A. Sharkey and N. Sharkey, "Children, the Elderly, and Interactive Robots," in IEEE Robotics & Automation Magazine, 2011, 18(1):32-38. DOI: 10.1109/MRA.2010.940151.

[27] Richard Cytowic, The Man Who Tasted Shapes. MIT Press, 2000. ISBN:9780262531528.

[28] Shuang Zhang, et al. The influence path of caregivers' positive aspects, expressed emotion and coping style on behavioral and psychological symptoms of dementia, Geriatric Nursing, 2022, 2022(44):143-150. DOI:10.1016/j.gerinurse.2022.01.013.

[29] Wolf K. Measuring facial expression of emotion. Dialogues in clinical neuroscience, 2015, 17(4): 457–462. DOI:10.31887/DCNS.2015.17.4/kwolf.

[30] Andrew G. Howard. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017. DOI:10.48550/arXiv.1704.04861

[31] Paul-Louis Pröve. An Introduction to different types of Convolutions in Deep Learning. Medium. Jul 23, 2017. https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d.

[32] Yoon Kim. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014, pp.1746–1751. DOI:10.3115/v1/D14-1181.

[33] Chollet François. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, (2016):1800-1807.

[34] Deepak Kumar Jain, Pourya Shamsolmoali, Paramjit Sehdev. Extended deep neural network for facial emotion recognition. Pattern Recognition Letters. 2019, (120):69-74. DOI:10.1016/j.patrec.2019.01.008.

[35] André Teixeira Lopes, et al. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. Pattern Recognition, 2017, (61):610-628. DOI:10.1016/j.patcog.2016.07.026.

[36] Arriaga O, Valdenegro-Toro M, Plöger P. Real-time convolutional neural networks for emotion and gender classification. Arxiv. 2017. DOI:10.48550/arXiv.1710.07557.

[37] Lorish, C. D., Maisiak, R. The Face Scale: a brief, nonverbal method for assessing patient mood. Arthritis and rheumatism. 1986, 29(7):906–909. DOI:10.1002/art.1780290714.

[38] Heerink M, Krose B, Evers V, Wielinga B. Assessing acceptance of assistive social agent

technology by older adults: the Almere model. Int J Soc Robot. 2010, (2):361–375.

[39] Caspar S, et al. Nonpharmacological Management of Behavioral and Psychological Symptoms of Dementia: What Works, in What Circumstances, and Why?. Innovation in aging. 2018, 2(1):igy001. DOI:10.1093/geroni/igy001.

[40] Li Y-Q, et al. Non-pharmacological interventions for behavioral and psychological symptoms of dementia: A systematic review and network meta-analysis protocol. Front. Psychiatry. 2022, 13:1039752. DOI: 10.3389/fpsyt.2022.1039752.

[41] Rachael E Jack, et al. Facial expressions of emotion are not culturally universal. PNAS. 2012, 109 (19):7241-7244. DOI:10.1073/pnas.1200155109.

[42] Hiroko Kase, et al. 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. 2019. Special Issue (10):S51-55. DOI:10.6890/IJGE.201910/SP.0008.

[43] Nomura N and Hashimoto T. Shinrigaku kenkyu: The Japanese journal of psychology. 2006, 77(1):32–39. DOI:10.4992/jjpsy.77.32.

[44] Tsai TH, Huang CC and Zhang KL. Design of hand gesture recognition system for human-computer interaction. Multimed Tools Appl. 2020, (79):5989–6007. DOI:10.1007/s11042-019-08274-w

[45] Sarthak Gupta, Siddhant Bagga, and Deepak Kumar Sharma. Hand Gesture Recognition for Human-Computer Interaction and Its Applications in Virtual Reality. Advanced Computational Intelligence Techniques for Virtual Reality in Healthcare. 2019, pp.85-105. DOI:10.1007/978-3-030-35252-3_5.

[46] Xiangyu Liu. Social Assistant Humanoid Robots for Dementia of the Elderly: a Scoping Review, Journal of Electronic Research and Application. 2020, 4(6):95-103. DOI:10.36922/ssr.v2i3.1001.

[47] Xiangyu Liu. Validation research on the application of Depthwise Separable Convolutional AI facial expression: recognition in non-pharmacological treatment of BPSD. Journal of Clinical and Nursing Research. 2021, 5(4). DOI:10.26689/jcnr.v5i4.2325.

[48] Berardina De Carolis, et al. 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.

[49] Akutsu Yasuko. Pepper's activity in the nursing care field. Medical Machinery. 2016, 86(4):393-396. (in Japanese).

[50] Yasuharu Den, et al. A Proper Approach to Japanese Morphological Analysis: Dictionary,

Model, and Evaluation. The Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA). 2008. http://www.lrec-conf.org/proceedings/lrec2008/pdf/258_paper.pdf.

[51] Aldebaran. Softbank Product Homepage. https://www.softbankrobotics.com/emea/en/pepper.

[52] Julien Martine, Jacques Jaussaud. Prolonging working life in Japan: Issues and practices for elderly employment in an aging society. Contemporary Japan. 2018, 30(2):224-242. DOI:10.1080/18692729.2018.1504530.

[53] Sugishita Morihiro, Hemmi Isao, and Takeuchi Tomoko. Reexamination of the Validity and Reliability of the Japanese Version of the Mini-Mental State Examination (MMSE-J). Japanese Journal of Cognitive Neuroscience. 2016, 18(3+4):168-183. DOI:10.11253/ninchishinkeikagaku.18.168.

[54] Shannon CE. A mathematical theory of communication. Bell System Tech J. 1948, 27(3):379–423. ISBN:0-252-72546-8.

[55] Lash A, et al. Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. Exp Aging Res. 2013, 39(3):235–253. DOI:10.1080/0361073X.2013.779175.

[56] Lodhi H, et al. Text classification using string kernels. J Mach Learn Res. 2002, 2:419–444. https://www.jmlr.org/papers/volume2/lodhi02a/lodhi02a.pdf.

[57] Suzuki M, et al. Accent sandhi estimation of Tokyo, dialect of Japanese using conditional random fields. IEICE Trans. Inf. &Syst. 2017, E100.D (4):655–661. DOI:10.1587/transinf.2016AWI0004.

[58] Shibuya N. Demystifying Entropy. San Francisco, US: A Medium Corporation; 2018. https://medium.com/activating-robotic-minds/demystifying-entropy-f2c3221e2550.

[59] Hidefumi Ohmura, Takuro Shibayama. Musical and Linguistic Structures Consisting of Pitch and Rhythm. The 30th Annual Conference of the Japanese Society for Artificial Intelligence. 2016. https://www.ai-gakkai.or.jp/jsai2016/webprogram/2016/pdf/398.pdf.

[60] Kitamura H, et al. Development of a direct behavioral observation assessment form for Behavioral and psychological symptoms of dementia (BPSD): Reliability and validity study. Journal of Higher Brain Functions. 30(4):510-522. (in Japanese) https://www.jstage.jst.go.jp/article/hbfr/30/4/30_4_510/_pdf.

[61] Kinsui Satoshi, Yukinori Takubo. A Study of "Yo," "Ne," and "Yonne" Based on Discourse Management Theory. Human-Machine Interaction through Speech.1998.pp.257-271. ISBN-10:4274078574.

[62] Clark H H. Using language. Cambridge University Press. 1996. ISBN: 9780511620539.

[63] Katagiri Y. Dialogue functions of Japanese sentence-final particles ‘Yo’ and ‘Ne’. J Pragmat. 2007, 39(7):1313–1323. DOI:10.1016/j.pragma.2007.02.013.

[64] Hayano, K. When (not) to claim epistemic independence: The use of ne and yone in Japanese conversation. East Asian Pragmatics. 2018. 2(2):163-193. DOI:10.1558/eap.34740.

[65] Nobutake Nomura, Tuskasa Hashimoto. Reminiscence, Identity Status and Well-being in Young Adults, The Japanese Journal of Personality. 2006. 15(1):20-32. DOI:10.2132/personality.15.20.

[66] Whiting D, et al. Effect of COVID-19 on BPSD severity and caregiver distress: Trend data from national dementia-specific behavior support programs in Australia. Alzheimer's & dementia: the journal of the Alzheimer's Association. 2021. 17(Suppl 12):e058454. DOI:10.1002/alz.058454.

[67] Kung TH, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023, 2(2): e0000198. DOI:10.1371/journal.pdig.0000198.

Appendix

A.1 MMSE scale

Mini-Mental State Examination (MMSE)検査シート

（医療機関が記入し、コピーを診療情報提供書に添付してください）

	質問内容	正誤	点数
1（5点）	今年は何年ですか 今の季節は何ですか 今日は何曜日ですか 今日は何月ですか 今日は何日ですか		
2（5点）	ここは何県ですか ここは何市ですか ここは何病院ですか ここは何階ですか ここは何地方ですか（例 関東地方）		
3（3点）	物品名３個（相互に無関係） 検者は物の名前を一秒間に一個ずつ言う。その後、被験者に繰り返させる。 正答一個につき１点を与える。３例全て言うまで繰り返す。（６回まで） 何回繰り返したかを記す。【 回】		
4（5点）	100から順に７を引く（５回まで）。 または「フジノヤマ」を逆唱させる		
5（3点）	３で提唱した物品名を再度復唱させる		
6（2点）	（時計を見せながら）これはなんですか （鉛筆を見せながら）これはなんですか		
7（1点）	次の文章を繰り返しさせる。 「みんなで力をあわせて綱を引きます。」		
8（3点）	（３段階の命令） 「右手にこの紙を持ってください」 「それを半分に折りたたんでください」 「それを私に渡してください」		
9（1点）	（次の文章を読んでその指示に従ってください。） 「目を閉じなさい」		
10（1点）	（何か文章を書いてください。）		
11（1点）	（次の図形を書いてください） 		
満点は30。カットオフポイント：23/24			

（森 悦郎他：神経疾患患者における日本語版 Mini-Mental State テストの有用性. 臨床心理学 1985; 1: 2-10）

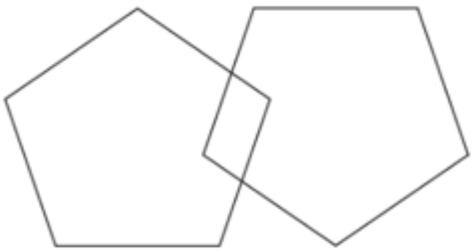
９．「この文を読んで、この通りにしてください」

「目（め）を閉（と）じてください」

１０．「この部分に何か文章を書いてください。どんな文章でもかまいません」

［　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　］

１１．「この図形を正確にそのまま書き写してください」



A.2 NPI-NH scale

日本語版 Neuropsychiatric Inventory
施設版(NPI-NH)

氏名＿＿＿＿＿＿＿＿　　評価日＿＿＿＿＿＿＿＿
IDNo.＿＿＿＿＿＿＿＿　　部屋番号＿＿＿＿＿＿＿＿

	N/A	頻度	重症度	頻度と重症度の積	負担度
妄想		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
幻覚		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
興奮		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
うつ		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
不安		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
多幸		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
無関心		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
脱抑制		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
易刺激性		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5
異常行動		0 1 2 3 4	0 1 2 3		0 1 2 3 4 5

情報提供者＿＿＿＿＿＿＿＿

職(看護婦、介護福祉士、ヘルパー、その他＿＿＿＿＿＿＿＿)

通常の勤務時間帯＿＿＿＿＿＿＿＿

情報提供者の介護者に関する知識 (非常によく知っている／毎日介護している;ある程度知っている／しばしば介護をしている;それほど知らない／薬を配るのみであったり、最小限の関わりしかなかったりする)

情報提供者のこの入所者を介護する1週間の勤務数＿＿＿＿＿＿＿＿

診断＿＿＿＿＿＿＿＿

治療内容＿＿＿＿＿＿＿＿

A.3 DBD scale

認知症行動傷害尺度 (Dementia Behavior Disturbance Scale : DBD)

次の1から28の項目について、次の0から4までの評価に従って記入してください。

0:全くない　1:ほとんどない　2:ときどきある　3:よくある　4:常にある

記入欄	項 目
	1　同じことを何度も何度も聞く
	2　よく物をなくしたり、置場所を間違えたり、隠したりしている
	3　日常的な物事に関心を示さない
	4　特別な理由がないのに夜中起き出す
	5　特別な根拠もないのに人に言いがかりをつける
	6　昼間、寝てばかりいる
	7　やたらに歩き回る
	8　同じ動作をいつまでも繰り返す
	9　口汚くののしる
	10　場違いあるいは季節に合わない不適切な服装をする
	11　不適切に泣いたり笑ったりする
	12　世話をされるのを拒否する
	13　明らかな理由なしに物を貯め込む
	14　落ちつきなくあるいは興奮してやたら手足を動かす
	15　引き出しやタンスの中身を全部だしてしまう
	16　夜中に家の中を歩き回る
	17　家の外に出ていってしまう
	18　食事を拒否する
	19　食べ過ぎる
	20　尿失禁する
	21　日中、目的なく屋外や屋内をうろつきまわる
	22　暴力を振るう(殴る、かみつく、引っかく、蹴る、唾をはきかける)
	23　理由もなく金切り声をあげる
	24　不適当な性的関係を持とうとする
	25　陰部を露出する
	26　衣服や器物を破ったり壊したりする
	27　大便を失禁する
	28　食物を投げる
	0点以外は異常

(溝口 環他:DBD スケールによる老年期痴呆患者の行動異常評価に関する研究. 日老雑誌 1993;30:835-840)