Context-Enhanced Human-Robot Interaction: Exploring the Role of System Interactivity and Multimodal Stimuli on the Engagement of People with Dementia

Yuan Feng[1,2] · Giulia Perugia[3,4] · Suihuai Yu[2] · Emilia I. Barakova[1] · Jun Hu[1] · G. W. Matthias Rauterberg[1]

Abstract

Engaging people with dementia (PWD) in meaningful activities is the key to promote their quality of life. Design towards a higher level of user engagement has been extensively studied within the human-computer interaction community, however, few extend to PWD. It is generally considered that increased richness of experiences can lead to enhanced engagement. Therefore, this paper explores the effects of rich interaction in terms of the role of system interactivity and multimodal stimuli by engaging participants in context-enhanced human-robot interaction activities. The interaction with a social robot was considered context-enhanced due to the additional responsive sensory feedback from an augmented reality display. A field study was conducted in a Dutch nursing home with 16 residents. The study followed a two by two mixed factorial design with one within-subject variable - multimodal stimuli - and one between-subject variable - system interactivity. A mixed method of video coding analysis and observational rating scales was adopted to assess user engagement comprehensively. Results disclose that when additional auditory modality was included besides the visual-tactile stimuli, participants had significantly higher scores on attitude, more positive behavioral engagement during activity, and a higher percentage of communications displayed. The multimodal stimuli also promoted social interaction between participants and the facilitator. The findings provide sufficient evidence regarding the significant role of multimodal stimuli in promoting PWD's engagement, which could be potentially used as a motivation strategy in future research to improve emotional aspects of activity-related engagement and social interaction with the human partner.

1 Introduction

Dementia, a neurodegenerative disease addressed by the World Health Organization and Alzheimer's Disease International as a public health priority [51]. It is not a part of normal aging and can erode people with dementia's (PWD) ability to perform daily tasks as they will gradually experience reduced cognitive ability, loss of memory, learning skills, language ability, and impaired affect regulation. With no existing cure in sight, the condition of the PWD can only get worse with the affected behaviours further exaggerated. The need for a high level of assistance, professional and inten-

✉ Yuan Feng
  Y.Feng@tue.nl

  Giulia Perugia
  giulia.perugia@it.uu.se

  Suihuai Yu
  caid@nwpu.edu.cn

  Emilia I. Barakova
  e.i.barakova@tue.nl

  Jun Hu
  j.hu@tue.nl

  G. W. Matthias Rauterberg
  g.w.m.rauterberg@tue.nl

1 Department of Industrial Design, Eindhoven University of Technology, Eindhoven, the Netherlands

2 Department of Industrial Design, Northwestern Polytechnical University, Xi'an, China

3 Human-Technology Interaction Group, Eindhoven University of Technology, Eindhoven, the Netherlands

4 Department of Information Technology, Uppsala University, Uppsala, Sweden

sive care means that most PWD are eventually admitted to long-term care (LTC) facilities where they can receive quality care. Such facilities can be efficient in meeting physical needs (e.g., hygiene, meals, place to live or medication use), but often fail to address psycho-social needs [26,62]. Consequently, the well-being of PWD in LTC facilities is hindered, as they spend most of their time alone, disengaged, have limited meaningful social interactions, and are exposed to inappropriate sensory stimulation (e.g., lack of sensory stimulation or over-stimulating by environmental factors). This prolonged lack of engagement in sensory stimulation, physical and social activities can further lead to accelerated disease development and worse living conditions.

Enhancing engagement is key for promoting the quality of life of PWD, especially for those living in residential homes [46,67]. It has been well-recognized in the literature that engagement in meaningful activities is associated with reduced challenging behaviors (e.g., agitation) [50], decreased psychological symptoms such as aggression, depression, and apathy [13,43,61], increased social connections [9,18], and improved positive emotions [45,64,74]. Therefore, in order to help PWD to live well with the disease after the diagnosis, it is essential to design and develop activities that could engage this special user group. The use of various kinds of interactive technologies to facilitate engagement of PWD is a rising field that is reshaping contemporary dementia care [49]. Despite the fact that most of the technology applications are still focusing on promoting self-independence, safety, or care practices and services of PWD [66], researchers are now aware that the priorities include not only attending activities that are important for independence but also for enjoyment, social interaction and relaxation [41]. Interactive system mediated psycho-social activities have huge potential in enabling motivation for active participation, providing sensory and physical stimulation, addressing social and emotional needs, and benefiting the mental health of PWD, as well as maintaining dimensions of human existence such as self-determination and dignity. From the design research perspective, it is therefore interesting to find out how could interactive system mediated activities be designed towards an increased level of engagement for PWD?

For decades, researchers within the human-computer interaction (HCI) community have been exploring how unique system features can influence user engagement. The framework of "Richness, Control and Engagement" proposed by Rozendaal [59] addressed the role of experienced richness and control in determining user engagement. The notion *richness* was described as "the range of possibilities afforded by an interactive medium in terms of perception and action", and were influenced by system features at sensorial level - the variety of external sensory stimulation; behavioral level - degree of various behavioural movements enabled; and mental level - curiosity and ambiguity through thought

process [57,58]. "Control" emphasizes the balance between personal experiences/skills and system provided challenges. The experienced richness is suggested by literature accumulated by a system afforded feature named "Interactivity", and the representational richness of a medium named "Vividness" [56]. Interactivity-which is central to interactive system design-has been well studied for decades and has been defined in many ways. The concept has many varied interpretations according to different perspectives. Interactivity defined by Steuer [63] combines both the possibilities of the system and the human action that is needed to bring about these possibilities. With more possibilities to manipulate the system in order to achieve higher goals, the interactivity is therefore increased. For an interaction experience, interactivity has the ability to influence the feeling of control and increase richness at behavioural level affecting the physicality of interaction, while presented sensory feedback could affect richness at a sensorial level, therefore, potentially contributing to enhanced user engagement.

However, few studies further extend this research of engagement to dementia users. For PWD with diminished cognitive and functional abilities, and impaired sensory information processing and integration skills, their perceptual and behavioral experience may vary from a general understanding of how user experiences were shaped. Therefore, in this presented paper, we took the research of experienced richness and how it influences user engagement to a specific target user group - PWD. Explicitly speaking, we investigate how system features in terms of system interactivity and multimodal presentations could impact engagement of PWD based on a designed activity of context-enhanced human-robot interaction (HRI). To achieve this objective, we first introduce the system and activity design of a robot-assisted interactive installation named *LiveNature*. The design aims to engage PWD in LTC in multisensory experience through rich interactions [21]. Next, we describe the field study conducted in a Dutch nursing home involving 16 residents with dementia. Participants were engaged in interaction sessions with varied levels of system interactivity and multimodal stimuli that were implemented through different configurations of the system design. Lastly, we present the knowledge acquired from this study and discuss how it could benefit future robotics research and dementia care. This paper contributes by revealing the relationship of experienced richness and engagement of dementia users, and providing new insights about the impact of multimodal stimuli and system interactivity on user engagement which will help to design interactive systems for PWD in LTC.

2 Context-Enhanced Human-Robot Interaction

The use of social robots in dementia care has been intensively investigated to optimize PWD's emotional and social well-being. Despite humanoid robots with appearance that resembles a human, robots with animal appearance were tested as reasonable substitutes of real animals and have demonstrated similar effects for evoking positive human emotions and motivating communications [7,47,70]. Within animal-like social robots research, most effort has been invested in gathering evidence to prove the effectiveness of interaction with off-the-shelf robots on promoting social engagement [35,60], supporting care activity [4], and regulating challenging behaviors such as anxiety, depression and agitation [10,25,48]. Other researchers looked into how to improve robotic designs to serve the emotional and mental needs of PWD better, and these works were well discussed with ethics reflections of robot use for elderly in general [24,42]. While social robotic studies have presented promising evidence in engaging dementia users in social activities, most researches on HRI with PWD were performed between a robot and a use/users alone, and only a few addresses how the context of interaction could potentially positively influence HRI experiences of PWD [20,28]. To our best knowledge, there has been limited research exploring HRI that incorporated contextual cues from a larger scale setting that attempts to engage PWD in a more sensory immersive experience with rich interaction possibilities [29].

Therefore, the research reported in this paper concerns an activity design of *LiveNature* that addresses our proposed notion of context-enhanced HRI [21]. *LiveNature* is an interactive system design aiming to connect residents to the outdoors through an indoor interactive experience due to their limited contact with real nature, especially during this COVID-prevalence times. On the one side, the system design consists of an augmented reality display mounted on the wall. On the other side, it consists of a sheep-like social robot, see Fig. 1. While the former attempts to provide immersive multi-sensory experience through dynamic media content and tangible augmentations, the latter strives to reinforced the engaging experience with rich interaction possibilities using enhanced tactile feedback and proximal embodied interaction through HRI [3]. The augmented reality display is a system unit containing an 87-inch ultra-high-definition display, a computer control system including sensors and actuators, and the tangible extension of the virtual content that enables simple pumping interactions. The robotic sheep was a prototype developed through re-programming a commercially available PLEO robot (a robotic dinosaur) using the Pleorb Development Kit (PrbDK). We disguised the PLEO's appearance to a lamb and equipped it with a furry textile and a soft stuffing material underneath. The PLEO was chosen due

to its well-developed behaviors that aim to provoke human emotions (e.g., happy gestures expressed using its head, neck, legs, and tail); and the tolerable level of mechanical sound caused by the motor, so that it is more likely to be perceived as an animal than a machine.

The interaction experience of *LiveNature* is context-enhanced in the following way: (1) the external sensory stimulation was provided not only from the robot but enhanced using visual-auditory sensory cues from a large display, which has a better chance of creating a more immersive sensory experience than the robot alone; (2) the media content simulates a life-like window overlooking a farm and provides a story narrative for easier facilitation and introduction of robot used at the beginning of a session. Here, the context refers to the simulated "closer to nature" experience, under which circumstances the HRI took place. This designed "context" aims to enhance the acceptance of the robot used as the literature suggests it is often challenging for PWD and their caregivers [73]; (3) the "context" also works as a periphery display. And by shifting between watching the display and interacting with the robotic sheep, users' interests might be potentially sustained.

The prototype of *LiveNature* was implemented in a Dutch residential care home. The activity was designed to provoke the playful experiences of *sensation*, *relaxation*, and *reminiscence* that were suggested to be suitable for the capacity of a larger audience of PWD regardless of the severity of their condition [2]. Since the original inclusive design process was undertaken with the Dutch elderly, see [21], the form and appearance design addresses several aspects that are familiar to this specific generation of users to trigger reminiscence and evoke positive emotional responses. Due to the fact that most residents had either grown up on a farm or had the farming experience, the display shows dynamic video content of a grass field with a heard of sheep to simulate a window outlook experience of typical Dutch farm scenery [38]. The nature media content and soundscapes were adopted to emulate nature-assisted therapy's soothing effects and avoid over-stimulation of senses in the LTC environment [33]. The robotic sheep works as a distributed tangible interface to interact with multisensory media content: when touching input from users was sensed, the robot behaves happy gestures through moving its head, neck, legs, and tail accompanied with lamb bleating; additionally, the content on display will change status from a more static to an active one (e.g., sheep herd becoming more alert and active, gathering in front of the display and curious about user's behavior).

3 Research Objective and Questions

Based on the above activity design of context-enhanced HRI, this study aims to gather a deeper understanding of how to

Fig. 1 The design of *LiveNature* that combines an augmented reality display mounted on the wall with a sheep-like social companion robot implemented in Vitalis nursing home as context-enhanced human robot interaction activity

better shape the engaging experience of PWD by exploring the role of system interactivity and multimodal stimuli in contributing to a successful interactive system design within the specific context of LTC. Therefore, the research questions related to the study aim are:

1. To what extent can different multimodal stimuli provided by system design influence the engagement of PWD living in the specific context within LTC?
2. To what extent can the level of system interactivity influence the engagement of PWD living in the specific context within LTC?
3. To what extent can the interaction effect of multimodal stimuli and level of system interactivity influence the engagement of PWD living in the specific context within LTC?

4 Method

4.1 Study Design and Setting

The field study was conducted within the real-life setting of an LTC for PWD with four experimental conditions and one control condition in total. The study design of experimental conditions followed a 2 by 2 mixed factorial experimental design with one within-subject variable - multimodal stimuli - and one between-subject variable - level of system interactivity. The system interactivity was considered increased when more interaction possibilities were enabled, and the level of multimodal stimuli was considered higher when external stimulation of more sensory channels was provided. The system configurations of *LiveNature* were modified to create different experimental and control conditions. Specifically, the levels of system interactivity (abbreviation as I) were divided according to whether the robotic sheep could be used as a tangible interface for triggering contextual interactions from the augmented reality display, and the levels of multimodal stimuli (abbreviation as M) were defined by whether auditory feedback was pre-

sented besides visual-tactile stimuli from both the robot and the display. In total, there were two levels of experimental conditions within each independent variable (named I1, I2, and M1, M2 respectively), and with the number increases, the level of independent variables increases. The experimental conditions with varying levels are presented with detailed descriptions in Table 1. In addition, we adopted a control condition for examining the group difference of engagement at baseline. During the control condition, participants were engaged in interaction with the augmented reality display only.

The prototype was situated in the hallway of a residential dementia care setting - Vitalis Kleinschalig Wonen (Vitalis for short), Eindhoven, the Netherlands. The public space was connected to private homes and living rooms so that the users could freely walk to. The experiment environment has large windows to the outside, receives sufficient sunlight, and with a controllable noise level, therefore ideal for visual-audio presentations of the study. Two seats were positioned in front of the display (for one-to-one interaction session of a participant and a facilitator) to create a comfortable atmosphere and accommodate wheelchair users. All experiment sessions were recorded with one primary camera (C1, a Microsoft Kinect camera installed right above the display facing directly towards the participants) and two supporting cameras (C2 - a GoPro camera placed on the left of the display, C3 - a digital camera place behind the participants). The setting of the experiment is shown in Fig. 2.

4.2 Participants

A total of 24 residents were recruited from the Vitalis nursing home. To estimate the required sample size of this study, we performed a priori statistical power analysis using the software package GPower (version 3.1.9.7) [19]. With effect size set at 0.40 (considered to be large according to Cohen's criteria), an alpha of 0.05, and power = 0.80, the projected sample size needed with this effect size is approximately N = 16 for this within-between interaction comparison. Thus, we recruited more than 16 participants at the beginning of

Table 1 Detailed descriptions of four experimental conditions with the level of multimodal stimuli as a within-subject variable and the level of system interactivity as a between-subject variable

|	Multimodal stimuli level 1 (M1): Visual-tactile stimuli provided	Multimodal stimuli level 2 (M2): Visual-auditory-tactile stimuli provided
	Condition M1I1	Condition M2I1
System interactivity level 1 (I1): The robotic sheep was disconnected from the system	The robotic sheep was turned Off and disconnected from the system; Visual content presented on display.	The robotic sheep was turned Off and disconnected from the system; Visual-auditory content was presented on display.
	Condition M1I2	Condition M2I2
System interactivity level 2 (I2): The robotic sheep was connected to the system	The robotic sheep was turned On with tactile-motion feedback; HRI triggers visual feedback from display.	the robotic sheep was turned On with tactile-motion-sound feedback; HRI triggers visual-auditory feedback from display.
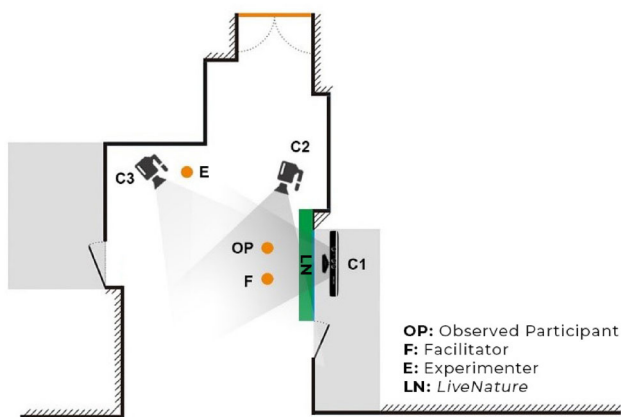
Fig. 2 Experiment settings with participants, stimuli and cameras

OP: Observed Participant
F: Facilitator
E: Experimenter
LN: *LiveNature*

the participant recruitment to make sure the sample size is adequate for the main objective of this study. We could not recruit more participants due to the limitation of the capacity of residents living in Vitalis, which is further discussed in the limitation section.

Inclusion criteria were: (1) a Mini-Mental State Examination (MMSE) score lower than 24 (25-30 was suggested as normal cognition, and below 24 as cognitive impairment); (2) signed informed consent of participants or their legal guardians. The exclusion criteria were: (1) acute visual or auditory impairment reported by the caregivers; (2) inability to sit, hold or interact with an interactive artifact. Twenty-one participants met the inclusion criteria and were therefore enrolled in the study. Participants were stratified according to their cognitive abilities and randomly assigned to 1 of 2 groups. The initial sample size decreased to 16 during the experiment period due to participants' death (n = 1), hospitalization (n = 1), and dropouts because of other reasons (n = 3). The final sample consisted of 16 participants (4 male, 12 female, $M = 85.2$, $SD = 4.8$, age range 78–92 years)

with group 1 consisting of seven participants and group 2 of nine participants (uneven number of participants are due to uneven dropouts). Detailed demographic information provided by the medical staff of participants is presented in Table 2. We ran a number of t-tests with the group as an independent variable and the socio-demographic and clinical characteristics of the group members as dependent variables. The results suggested no significant differences between the two groups on each characteristic (see also Table 2). Each participant took part in three sessions in total (including one control condition and two experimental conditions) with one session per week. For instance, group 1 would participate in the control condition, condition M1I1, and condition M2I1; and for group 2, the control condition, condition M1I2, and condition M2I2. The participation order was randomly chosen from all six possibilities of the permutation of three conditions to control counterbalancing effects and assigned to each participant before the whole sessions started.

4.3 Measures

Evaluation of engagement with measures that are reliable, valid, and robust is essential for designing interactive systems from users' perspective. The notion of engagement is challenging to capture, and it is more challenging for PWD due to the accompanied cognitive, functional, and language impairments of the disease. This study adopted a mixed method of video coding analysis and observational rating scales for a comprehensive assessment of PWD's engagement. Two types of measures were adopted using different data collections, including: (1) video and audio recordings of the whole experiment sessions of all experimental conditions were recorded for video coding analysis using an observational video coding scheme - Ethographic and Laban-Inspired Coding System of Engagement (ELICSE) [52,55]; (2) rating data of all sessions of both control and

Table 2 Socio-demographic characteristics of the participants

Characteristics	G1 (n = 7)	G2 (n = 9)	p-value
Age (years)			.33
Mean (SD)	86.6 (4.2)	84.1 (5.2)	
Age range	80–92	78–92	
Gender, n (%)			.79
Female	5 (71.4)	7 (77.8)	
Male	2 (28.6)	2 (22.2)	
Marital status, n (%)			.60
Single/divorced	1 (14.3)	1 (11.1)	
Married/partner	4 (57.1)	4 (44.4)	
Widowed	2 (28.6)	4 (44.4)	
Type of dementia, n (%)			.72
Alzheimer's dementia	2 (28.6)	3 (33.3)	
Vascular dementia	1 (14.3)	2 (22.2)	
Other/Mixed dementia	4 (57.1)	4 (44.4)	
MMSE score			.48
Mean (SD)	14 (5.3)	11.3 (8.3)	
Score range	8–22	0–23	
Dementia severity			.86
Mild	1 (14.3)	1 (11.1)	
Middle	2 (28.6)	3 (33.3)	
Middle to Severe	3 (42.9)	2 (22.2)	
Severe	1 (14.3)	3 (33.3)	
Mobility, n (%)			.72
Use wheelchair	3 (42.9)	3 (33.3)	
Use stroller	2 (28.6)	4 (44.4)	
Use none	2 (28.6)	2 (22.2)	

Abbreviations, G1 - group 1; G2 - group 2

experimental conditions were collected using the scale of Observational Measurement of Engagement (OME) [14], the Observed Emotional Rating Scale (OERS) [40], and the Engagement of a Person with Dementia Scale (EPWDS) [37]. The interaction-triggered user engagement (short-term engagement) was assessed using OME, EPWDS, and video analysis based on ELICSE coding scheme, while the affective states of the participants were measured through OERS.

A trained research assistant who was blinded to the study's objectives completed the video coding analysis. Rating scales OME and OERS were completed through direct observation on-site by a facilitator, while EPWDS was rated by the same research assistant using videos for indirectly observation-based ratings. The EPWDS was rated based on off-site video recordings due to two reasons: 1) practical time limitation between arranged sessions; and 2) the EPWDS was developed based on a previous video coding tool named VC-IOE [36] and were originally evaluated using videos materials, see [37].

4.3.1 The ELICSE Coding Scheme for Assessing Engagement of Dementia

The ELICSE coding scheme was developed by Perugia et al. [55]. It aims to measure engagement in PWD through observational behaviors. The coding system was built based on the qualitative analysis of body movements to estimate engagement in activities and social interactions (e.g., direct manipulation using hands when playing puzzles indicates that participants are engaging with the game), and the resulting ethograms were structured based on Laban Movement Analysis [39,52]. The assessment of the intensity of engagement is gauged by observing the body/facial configurations of the person with dementia during the activity and associating them with an engagement score. The coding scheme is composed by *Behaviors* and *Modifiers*. Of which, the *Behaviors* identified in ELICSE measure changes in the direction of attention, and the *Modifiers* define whether such behaviors are associated with affective nuance.

The original coding scheme, as in [53], encompasses three behavioral modalities involving three different body

parts respectively: the *Head*, the *Torso* and the *Arms/Hands*. In order to apply the ELICSE to our specific study, we adapted the original coding scheme considering body portion involvement under the specific context of interaction with *LiveNature*. Three pilot tests were carried out with three random participants to see how residents interacted with the designed interactive system to guide and determine the final coding scheme. Based on the pilot test, we employed two modalities *Head* and *Arms/Hands Behaviors* from the original ELICSE coding scheme and removed the *Torso Behaviors*. As preliminary observations indicated, those participants in their later stages of the disease (or in the wheelchair) had few torso movements (i.e., torso position changes, e.g., leaning forward to show more engaged), in addition to the selected behavioral modalities, we used an additional cue - *Conversations* in the final coding scheme. The verbal behaviors are congruent with bodily behaviors and also fit the constructs by demonstrating attention focus through conversational counterpart and affective nuance through the content of verbal expressions. They have the potential to compensate disorders with facial expression or mobility deterioration, hence providing more comprehensive measures of observable facets of engagement.

The adapted ELICSE coding scheme was constructed by three main components: (1) bodily parts that express behaviors involved in engagement (e.g., *Head Behaviors*, *Arms/Hands Behaviors* and *Conversations*); (2) a cluster of behaviors which all former body parts share the same focus to demonstrate their focus of attention (e.g., towards *Facilitator*, *Augmented Reality Display*, *Robotic Sheep*, or *None of the Target*); and (3) modifiers added on former behaviors that express a positive, neutral, or negative affective nuance (e.g., *Positive*, *Neutral*, and *Negative Signs of Affection*). The final coding scheme used in the analysis is presented in Table 3.

4.3.2 Observational Rating Scales for Assessing Engagement and Affective States

Three observational rating scales with different emphasis in terms of engagement evaluation were employed in this study. OME, a seven-point Likert scale developed by Cohen-Mansfield and colleagues based on the Comprehensive Process Model of Engagement [16] was adopted as a direct observation measure of activity engagement in PWD. A short version of OME containing two main categories that reflected the user engagement in terms of *Attention* and *Attitude* was used in this evaluation. *Attention* measures the amount of attention the participant is paying to stimulus during the engagement. It can be behavioral (e.g., stroking the robotic sheep even if looking away), visual (e.g., staring at the robot even if not interacting with it), or conversational (e.g., talking about stimulus-related experiences). *Attitude*

measures the amount of excitement/expressiveness toward the stimulus (e.g., smiling, frowning, excitement in voice). Each category measures engagement through two subcategories - *Most of the Time* and *Highest Level*. The former reflects the attention and attitude towards the stimulus in an average situation, and the latter represents the highest level of attention (if the participant is very attentive for a little while and somewhat attentive most of the time, then mark a 2 for *Most of the Time* and a 4 for *Highest Level* according to the manual of OME [14]).

OERS is another generally used observation-based Likert scale that aims to measure the extent of emotion expressions during a session. This five-point Likert scale has descriptive indicators for five affective states: *Pleasure*, *Anger*, *Anxiety/Fear*, *Sadness*, and *General Alertness* [40]. We used two of the items *Pleasure* and *General Alertness* in this study. OERS was rated based on the extent of each affect expressed towards both the stimulus and human partners (if any applicable). A higher score indicates a greater display of a particular effect.

In addition, EPWDS, a five-point Likert scale, was also adopted for evaluating user engagement within LTC setting [37]. Differentiated from OME, which mainly focuses on activity participation (engagement with the stimulus), EPWDS emphasizes the social interaction of PWD as well. The scale could compute an overall score to represent engagement states that could be easily compared across different conditions. This 10-item scale measures five dimensions of engagement: *Affective*, *Visual*, *Verbal*, *Behavioral* and *Social Engagement*. Each dimension was assessed separately using a positive and a negative sub-scale and interpreted collectively to provide an overall impression of all facets of engagement. Item 2, 4, 6, 8, and 10 are reverse scored items, meaning after scoring is completed, the scored numerical scoring needs to be reversed to calculate the overall number that measurement engagement. Each item indicates the extent to which the rater agrees or disagrees with the statement ("strongly disagree" = 1, "strongly agree" = 5). The total score ranges from 10-50 if all items across the scale are rated. A higher total score indicates higher positive engagement exhibited.

4.4 Procedure

An experimenter and a facilitator were on site to ensure the proper facilitation of study sessions. The experimenter's role was to (1) configure the interactive system design as required by each condition; (2) supervise the study procedures and provide explanations when necessary; (3) manage all the recording devices for proper data collection. The same facilitator facilitated all the study sessions (both experimental and control conditions), and was trained extensively through pre-experiment presentations, written guidelines, and received

Table 3 The adapted ELICSE coding scheme used for scoring video recordings of all experimental sessions

Behaviors	Modifiers
Head (Gaze) Behaviors	Signs of Affection
Gaze toward Facilitator (Gaze_F)	-With positive signs of affection (_Pos)
Gaze toward Augmented Reality Display (Gaze_ARD)	-With neutral of affection
Gaze toward the Robotic Sheep (Gaze_RS)	-With negative signs of affection (_Neg)
None of the target gaze behaviors (Gaze_None)	
Arms/Hands Behaviors	Quality of Reach Out
Reach out Facilitator (Reach_F)	-Warmly reach out (_Pos)
Reach out Augmented Reality Display (Reach_ARD)	-Neutrally reach out
Reach out/Manipulate the Robotic Sheep (Reach_RS)	-Negatively reach out (_Neg)
None of the target hand gestures (Reach_None)	
Conversations	Quality of Conversations
Talk to the Facilitator (Talk_F)	-Positive verbal engagement with stimulus or the facilitator (_Pos)
Talk to the Robotic Sheep/Sheep on the Screen (Talk_Sheep)	-Neutral verbal engagement
Talk to themselves (Talk_Self)	-Negative verbal engagement with stimulus or the facilitator (_Neg)
Not understandable conversations (Talk_None)	
Silence (Talk_Sil)	

Behaviors marked in italic style are assigned with modifiers (i.e., positive, neutral, negative nuance). The "stimulus" here refers to both the augmented reality display and the robotic sheep

regular personal supervision throughout the study to get familiar with the study procedures. The study was arranged during non-planned activity times (i.e., 10:00–12:30 a.m. and 14:00–16:00) to accommodate daily care schedules and to control the high behavioral time of the day (e.g., the Sundowning effect, which describes the challenging behaviors that often appears before dinner time). Individual sessions lasted up to 20 minutes, long enough for explorations and short enough to not be interrupted by nursing care or visitors.

Pre-interaction session: Demographic data were collected by the facilitator before interaction sessions. And all recruited participants were asked to fill in the MMSE with the help of the facilitator. Before each interaction session started, the facilitator was instructed first to introduce the experiment's intention to participants and spend some time together with the participant to get acquainted. Participants were then invited for a one-on-one interaction session with the consideration of their wishes and mood. Upon participant's agreement, the facilitator guided him/her, walked to where the study took place and sat in front of the display. In the meantime, the experimenter prepared the setting according to the conditions designed and then introduced and brought the robotic sheep to the participant once he/she arrived (if the condition required the robot). Afterward, the facilitator explained how the system could be interacted with and entertained the participant.

During the interaction session: After the brief introduction, the facilitator switched on the audio recorder and gave the experimenter a sign to imply the session started. The experimenter then turned on all three cameras to record the session. The facilitator facilitated the interaction with verbal encouragement until participants started to lose interest and focus, intended to leave, or reached the maximum time limitations. The facilitator was instructed to try to be inconspicuous while interacting, let the participants freely explore the system design, and encourage engagement when needed.

Post-interaction session: Once the sessions ended, the facilitator gave an ending sign to the experimenter so that all video/audio recordings were then turned off. The experimenter retrieved the robotic sheep and thanked the participant for his/her participation. The facilitator then accompanied the participant back to their living/private rooms and came back to complete the OME and OERS.

4.5 Data Analysis

The video coding analysis of ELICSE was completed using Noldus Observer XT 14.2 software. IBM SPSS Statistics Version 25 was used for data entry and statistical computations. There was no missing data as all 16 participants finished all experimental sessions. The critical *p*-value was set at 0.05 (= 5% alpha error). For inter-rater reliability (IRR), a second-rater (different from the facilitator or the research assistant who completed the rating of EPWDS and video coding analysis) rated and coded part of the sessions (40%, 13 out of 32 sessions, randomly selected from all experimental sessions). IRR of video coding analysis was calculated

using Observer XT (i.e., Reliability Analysis) with Cohen's kappa statistic [11]. When calculating IRR, the Observer XT software takes both the matching of scored behaviors by two coders and the overlap of time into consideration. We utilized the 'Frequency/sequence' method of comparison and set 3 seconds tolerance for reliability analysis. The IRR result of 13 paired sessions ranged from a minimum Kappa of 0.68 to a maximum Kappa of 0.90 with an average of Kappas 0.82. Moreover, the IRR of rating scales was calculated using Cohen's Kappa by SPSS. According to Fleiss [22], Kappa value between 0.40-0.60 was considered a fair agreement, between 0.60-0.75, a good agreement, above 0.75 an excellent agreement. Overall, the IRR for all rating items was between good and excellent, ranging from 0.61 to 0.78.

4.5.1 Video Coding Analysis Using ELICSE

Coding Procedures. Initially, video recordings from all three cameras and audio recordings of each session were synchronized to have the same starting and ending point. The synchronization of videos was achieved by editing the video and audio files using Adobe Premiere CC to the same length. A total of 32 video/audio-recorded sessions with a total duration of 5.8 hours were annotated using Observer XT. Three pilot sessions were randomly selected and used for discussing discrepancies of video annotation together with the rater (i.e., the trained student assistant). Before scoring the behaviors of a session, the rater was instructed to watch the whole video for a general overview, then code each behavior group (*Head Behaviors*, *Arms/hands Behaviors*, and *Conversations*) separately. Within behavior groups, each cluster of behaviors was scored as mutually exclusive with a continuous sampling technique. The non-verbal behaviors were scored mainly using the video footages from the primary camera - C1 - as they had the clearest view of facial expressions and body movements; while the verbal behaviors (*Conversations*) were scored using the audio recordings as they provided a higher technical quality. When coding analysis of all sessions completed, the absolute duration and percentage duration of each scored behavior and modifier was then exported for further data aggregation and pattern examinations.

Data Aggregation. As suggested by the previous work [52, 55], the observable facet of engagement measured through ELICSE is composed of two essential components: *Attention* and *Valence*. The scored *Behaviors* of ELICSE are associated with the component *Attention* (regardless of attentive or non attentive expressed) and the scored *Modifiers* are associated with the component *Valence* (regardless of positive, neutral, or negative valence expressed). In order to properly interpret the data collection of video coding analysis, we aggregated relevant scored values to represent the extent to which the user is engaged with the activity.

Table 4 Data Aggregation of scored behaviors and modifiers of ELICSE coding scheme

Aggregated Items	Data Aggregation Computation
Gaze_LN	Gaze_ARD + Gaze_RS
PosGaze_LN	Gaze_ARD_Pos + Gaze_RS_Pos
Reach_LN	Reach_ARD + Reach_RS
PosReach_LN	Reach_ARD_Pos + Reach_RS_Pos
TalkAct	Talk_F + Talk_Sheep + Talk_Self
PosTalkAct	Talk_F_Pos + Talk_Sheep_Pos + Talk_Self_Pos
NegTalkAct	Talk_F_Neg + Talk_Sheep_Neg + Talk_Self_Neg

Therefore, the non-verbal behaviors in ELICSE that are relevant to this engagement study (i.e., attention focus directed towards the augmented reality display and robotic sheep) were aggregated into items: *Gaze toward Live-Nature* (Gaze_LN) and *Reach out/Manipulate LiveNature* (Reach_LN). The verbal behaviors during the interaction sessions (i.e., scored items except *Not understandable conversations* or *Silence*) were aggregated into *Talk Activity* (TalkAct) to represent verbal engagement during a session. Similarly, the modifiers with the positive nuance of each category that are engagement related (i.e., positive valence directed towards the augmented reality display and robotic sheep) were aggregated into *Gaze toward LiveNature with positive signs of affection* (PosGaze_LN), *Warmly reach out/manipulate LiveNature* (PosReach_LN), and *Talk Activity with positive verbal engagement with the stimulus or the facilitator* (PosTalkAct) accordingly and the modifier with the negative valence of *Quality of conversations* (i.e., *Negative verbal engagement with the stimulus or the facilitator*) was aggregated into *Talk Activity with negative verbal engagement with the stimulus or the facilitator* (NegTalkAct). The reason for not including aggregated items of *Gaze toward LiveNature with negative signs of affection* (NegGaze_LN), *Negatively reach out/manipulate LiveNature* (NegReach_LN) was due to a very low occurrence of such behaviors during the video scoring procedure. For an overview of the data aggregation computation, see Table 4. A higher computed value of a certain aggregated item indicates a higher level of engagement or affective states for that specific category.

4.6 Ethical Considerations

This study was approved by the Board of Vitalis Woon-Zorg Groep care center, where written informed consent was obtained from participants or their legal guardians if participants are no longer capable of giving informed consent any more. The principal investigator contacted the nursing home to hold a family meeting prior to the experiment with legally

authorized representatives of residents for presenting all relevant information regarding the experiment, signing informed consent, and residents' rights to refuse to participate during any time. Written descriptions of the proposed research were emailed to unattended legal representatives. The research was permitted and conducted in accordance with the requirements of the Eindhoven University of Technology. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

5 Results

5.1 Manipulation Check for Baseline Control

To ascertain that the participants allocated to the two groups did not differ in user engagement at baseline, we performed independent sample t-tests on all rating scale items of the OME, OERS, and EPWDS gauged after the control sessions between the two groups. Data collected using three rating scales were summarized using the means and Standard Deviations (SDs), see Table 5. The results indicated that there was no significant difference on all rating items except *Attention Highest Level* (Atten_H) $t(14) = 2.357$, $p = .034$. Nevertheless, the item *Attention Highest Level* evaluates participants' highest level of attention during an interaction session. And since the *Attention Most of the Time* is not significantly different between the two groups, we considered that the participant allocation would not bias our further statistical analysis regarding the main research questions. However, we examine further statistical analyses of the item Atten_H with caution.

5.2 Effects of System Interactivity and Multimodal Stimuli on Engagement

To answer the main research questions proposed in sect. 3, statistical analyses were performed on all aggregated items of the ELICSE and rating scale items of the OME, OERS, and EPWDS. The Bonferroni corrections were used to avoid alpha inflation. The partial eta-squared was used for reporting the effect size due to the limited sample size. Suggested norms for partial eta-squared according to Cohen's guidelines are ≤ 0.01 is considered small, ≈ 0.06 as medium, and ≥ 0.14 as large [12].

5.2.1 Results of Video Coding Analysis Using ELICSE

The means and SDs of the length of the total duration of a session (*Total Duration*), aggregated items of ELICSE using *Absolute Duration*, and aggregated items of ELICSE

using *Percentage Duration* (i.e., calculated using absolute duration/length of the total duration of a session) were summarized in Table 6. Regards the collected data using *Absolute Duration*, we performed a multivariate analysis of variances with repeated measurements and adopted the total duration of a session as a co-variable. The level of system interactivity was used as a between-subject factor, and the multimodal stimuli presented was considered the within-subject factor. The results revealed a significant main effect of multimodal stimuli level on item PosReach_LN (i.e., warmly reach out the installation *LiveNature* including the augmented reality display and the robotic sheep) $F(1, 14) = 5.719$, $p = .031$, $\eta^2 = .290$, shown in Table 6. The above significant result indicates that participants with more sensory modalities engaged during the study showed significantly higher positive behavioral engagement in terms of warmly petting, touching, or playing behaviors with both the robotic sheep and augmented reality display. For collected data using percentage duration, the mixed factorial analysis of variance (ANOVA) tests showed a significant main effect of multimodal stimuli level on item TalkAct (i.e., the verbal expressions during the session) $F(1, 14) = 4.720$, $p = .047$, $\eta^2 = .252$, meaning the percentage of time that participants were engaged in verbal communications were significantly higher when stimuli with more sensory modalities were presented. We did not find any significant main effect on level of system interactivity nor interaction effects on items Gaze_LN, PosGaze_LN, Reach_LN, PosReach_LN, TalkAct, PosTalkAct, and NegTalkAct (see Table 6).

5.2.2 Results of Observational Rating Scales

We performed mixed factorial analysis of variance (ANOVA) using the level of system interactivity as a between-subject factor and the level of multimodal stimuli as a within-subject factor on all rating scale items. The results show significant main effects of multimodal stimuli level on *Attitude Most of Time* (Atti_M) of OME, $F(1, 14) = 7.574$, $p = .016$, $\eta^2 = .351$, *Visual Engagement* (Vis_E) of EPWDS, $F(1, 14) = 8.113$, $p = .013$, $\eta^2 = .367$, *Social Engagement* (Soc_E) of EPWDS, $F(1, 14) = 5.011$, $p = .042$, $\eta^2 = .264$, and *Overall Engagement* (Eng_Sum) of EPWDS, $F(1, 14) = 5.250$, $p = .038$, $\eta^2 = .273$, indicating that the attitude, visual engagement, social engagement and overall engagement of PWD significantly improve when more sensory modalities are provided by the interactive system design. We did not find any main effect on system interactivity or interaction effect. All outputs of the ANOVA analyses with relevant descriptive statistics and critical p-values were presented in Table 7.

Table 5 Independent sample *t*-tests on rating items from OME, OERS, and EPWDS of control condition to disclose whether there were significant difference of engagement at baseline between two group of participants

Items	Control Condition Mean (SD)		*p*-value
	G1 (N=7)	G2 (N=9)	
OME			
Atten_M	5.29 (.76)	4.78 (1.09)	.313
Atten_H	6.14 (.90)	5.22 (.67)	**.034**
Atti_M	4.57 (1.13)	4.89 (1.05)	.572
Atti_H	4.86 (1.07)	5.44 (.88)	.248
OERS			
Pleasure	2.29 (1.11)	2.11 (.60)	.693
Alertness	4.14 (9.90)	3.56 (1.01)	.248
EPWDS			
Aff_E	8.43 (2.07)	8.00 (1.58)	.645
Vis_E	7.57 (1.90)	7.11 (1.27)	.570
Ver_E	7.71 (.95)	7.67 (1.87)	.952
Beh_E	6.57 (.79)	6.67 (.87)	.824
Soc_E	6.43 (1.13)	6.22 (.67)	.655
Eng_Sum	36.71 (5.82)	35.67 (4.80)	.699

Significance in bold. Abbreviations: Atten_M - attention most of the time; Atten_H - attention highest level; Atti_M - attitude most of the time; Atti_H - attitude highest level; Aff_E - affective engagement; Vis_E - visual engagement; Ver_E - verbal engagement; Beh_E - behavioral engagement; Soc_E - social engagement; Eng_Sum - overall engagement

Table 6 Results of main and interaction effects of levels of system interactivity and multimodal stimuli on all items of ELICSE coding scheme of four experimental conditions

ELICSE Items	Experimental Conditions Mean (SD)				MS		I		Interaction	
	M1I1 (N=7)	M2I1 (N=7)	M1I2 (N=9)	M2I2 (N=9)	*p*	η^2	*p*	η^2	*p*	η^2
Total Duration	473.00 (83.19)	831.71 (112.01)	731.56 (73.37)	657.56 (98.78)	.177	.126	.628	.017	**.049**	.250
Absolute Duration										
Gaze_LN	243.04 (146.52)	375.86 (208.12)	346.04 (176.94)	325.99 (239.62)	.256	.091	.767	.007	.130	.156
PosGaze_LN	11.96 (22.94)	19.54 (25.76)	29.41 (30.95)	44.67 (87.32)	.506	.032	.257	.091	.794	.005
Reach_LN	234.34 (256.59)	382.09 (311.47)	360.99 (242.73)	361.79 (314.27)	.397	.052	.649	.015	.402	.051
PosReach_LN	26.96 (22.70)	71.48 (59.09)	107.01 (133.49)	135.75 (184.81)	**.031**	.290	.259	.090	.614	.019
TalkAct	334.30 (294.77)	692.13 (408.57)	544.313 (265.39)	539.45 (258.08)	.117	.166	.802	.005	.108	.174
PosTalkAct	69.28 (96.93)	61.94 (100.46)	30.00 (62.34)	62.07 (101.74)	.484	.036	.651	.015	.271	.086
NegTalkAct	34.06 (50.74)	82.03 (72.15)	29.44 (67.31)	42.70 (50.43)	.103	.178	.396	.052	.340	.065
Percentage Duration										
Gaze_LN	54.09 (20.74)	47.27 (17.56)	51.10 (26.41)	48.37 (28.44)	.351	.062	.934	.001	.686	.012
PosGaze_LN	2.70 (3.92)	3.60 (6.52)	4.34 (4.34)	4.89 (7.76)	.693	.011	.547	.026	.927	.001
Reach_LN	44.89 (29.53)	55.39 (33.50)	52.68 (33.22)	48.99 (37.02)	.689	.012	.964	.000	.410	.049
PosReach_LN	9.79 (11.14)	11.56 (13.67)	16.25 (19.96)	17.00 (18.04)	.691	.012	.455	.041	.872	.002
TalkAct	66.60 (31.92)	77.04 (29.76)	71.69 (19.34)	83.64 (19.52)	**.047**	.252	.619	.018	.886	.002
PosTalkAct	20.68 (22.66)	11.67 (24.09)	4.48 (9.01)	6.85 (10.09)	.438	.043	.180	.125	.193	.118
NegTalkAct	6.88 (7.89)	8.37 (6.18)	3.02 (5.95)	7.10 (9.26)	.179	.125	.440	.043	.522	.030

Significance in bold. Abbreviations, MS - level of multimodal stimuli; I - level of system interactivity; Interaction - interaction of level of multimodal stimuli and system interactivity

Table 7 Results of main and interaction effects of levels of system interactivity and multimodal stimuli on rating scale items of OME, OERS, and EPWDS among experimental conditions

Scale Items	Experimental Conditions Mean (SD)				MS		I		Interaction	
	M1I1 (N=7)	M2I1 (N=7)	M1I2 (N=9)	M2I2 (N=9)	p	η^2	p	η^2	p	η^2
OME										
Atten_M	5.14 (0.90)	5.86 (0.69)	5.33 (0.87)	5.56 (1.01)	.074	.210	.884	.002	.328	.068
Atten_H	5.86 (1.07)	6.43 (0.54)	6.00 (0.71)	6.22 (0.83)	.201	.114	.911	.001	.564	.024
Atti_M	4.29 (1.38)	4.86 (0.69)	4.22 (0.97)	5.33 (1.23)	**.016**	.351	.662	.014	.392	.053
Atti_H	5.14 (1.57)	5.71 (0.95)	5.22 (1.09)	5.78 (1.09)	.109	.173	.888	.001	.981	.000
OERS										
Pleasure	2.14 (0.69)	2.14 (0.38)	2.22 (0.67)	2.33 (1.00)	.806	.004	.658	.014	.806	.004
Alertness	4.29 (0.76)	4.43 (0.79)	4.11 (0.93)	4.44 (0.73)	.301	.076	.819	.004	.674	.013
EPWDS										
Aff_E	7.71 (1.70)	8.71 (1.38)	7.44 (1.88)	8.44 (1.67)	.063	.226	.071	.011	1.000	.000
Vis_E	8.00 (2.08)	9.00 (1.53)	7.56 (2.35)	8.78 (1.72)	**.013**	.367	.719	.010	.780	.006
Ver_E	7.71 (2.56)	8.00 (2.24)	7.56 (1.74)	8.11 (1.76)	.432	.045	.979	.000	.799	.005
Beh_E	7.86 (2.85)	8.43 (1.51)	7.11 (2.85)	8.67 (1.32)	.072	.213	.802	.005	.383	.055
Soc_E	6.57 (2.15)	7.57 (1.27)	6.56 (1.67)	7.56 (1.81)	**.042**	.264	.984	.000	1.000	.000
Eng_Sum	37.86 (10.30)	41.71 (7.36)	36.22 (9.19)	41.56 (7.83)	**.038**	.273	.822	.004	.718	.010

Significance in bold. Abbreviations: MS - level of multimodal stimuli; I - level of system interactivity; Interaction - interaction of level of multimodal stimuli and system interactivity; Atten_M - attention most of the time; Atten_H - attention highest level; Atti_M - attitude most of the time; Atti_H - attitude highest level; Aff_E - affective engagement; Vis_E - visual engagement; Ver_E - verbal engagement; Beh_E - behavioral engagement; Soc_E - social engagement; Eng_Sum - overall engagement

6 Discussion

In this section, we first discuss the results presented above and emphasize the most interesting findings. We then discuss the mixed methodology use of ELICSE besides the golden standards of observational rating scales. In addition, we summarize the implications that contribute to future HRI studies, robotic research and development. Limitations and future works are also addressed.

6.1 Discussion on Experimental Effects

6.1.1 Contributions of Multimodal Stimuli on Promoting Engagement, Attitude and Communications

In general, the results obtained through the mixed methodology use of the ELICSE and the rating scales indicate that the level of multimodal stimuli had a significant impact on overall user engagement (according to the result of Eng_Sum of EPWDS), attitude (Atti_M of OME), valence (PosReach_LN of ELICSE), verbal communications (TalkAct of ELICSE), visual engagement (Vis_E of EPWDS), and social engagement (Soc_E of EPWDS), see Tables 6 and 7. Participants demonstrated significantly more positive behavioral engagement and a higher percentage of duration of verbal expressions when auditory stimuli were presented on the basis of visual-tactile feedback during the study. Besides, the atti-

tude towards the provided activity (during most of the time), the visual engagement, and social engagement were higher when more sensory modalities were involved in the interaction sessions. The above-mentioned findings are in line with previous research stating that everyday sound (i.e., nature soundscape and animal sound in this study) has promising benefits in dementia care as it can stimulate meaningful connections with past memories as well as among interpersonal human interactions [31]. In this specific study setting, adding content-relevant auditory feedback worked as a proactive strategy for facilitating verbal communications and positive affect display even if participants' visual and tangible/tactile sensory modalities were already engaged. As most recent design research that targets PWD with advanced stages tends to emphases on tangible/textile interaction [5,34,68], incorporating sound together with touch explorations could be one promising answer for positive engaging experience design of PWD.

Although we have exhibited significant results of multimodal stimuli level on Atti_M and PosReach_LN, other items that also accessed users' affect (i.e., PosGaze_LN of ELICSE, *Pleasure* of OERS, and Affe_E of EPWDS) did not reveal any statistical significance. To further understand this, we need to know that although many items seem to be conceptually overlapped, each assessment tool has its emphasis. And these are reflected in two aspects: (1) whether the focus of the assessment was on activity-related engagement only or

activity and interpersonal social engagement (human-human interaction) as an entity; and (2) whether it was accessed based mainly on one dimension of facial, behavioral and verbal affective expression, or a combined interpretation of all above.

Specifically, according to the manual of OME, this measure was developed to assess user engagement with the provided stimulus/activity. Item *Attitude* of OME was rated based on the amount of excitement/expressiveness toward stimulus/activity (e.g., smiling, frowning, energy, excitement in voice), and assessed through a comprehensive interpretation of facial expressions, verbal expressions, and behavioral manipulations in combined. On the other hand, scales of OERS and EPWDS view the interaction with the stimulus/activity and human partner as an entity. Item *Pleasure* of OERS was rated based on intensity reflected by the duration of pleasure expressions displayed when engaged with both the provided activity and the facilitator. The pleasure expressions were defined by showing signs such as laughing, smiling, singing, kissing, or rapport behaviors with another human. And item Affe_E of EPWDS was rated based on to which extent the rater agreed with two statements: one positive according to [37] - "Displays positive affect such as pleasure, contentment or excitement (e.g., smile, laughing, delight, joy, interest and /or enthusiasm)"; and one negative - "Display negative affect such as apathy, anger, anxiety, fear, or sadness (e.g., disinterest, distressed, restless, repetitive rubbing of limbs or torso, repeated movement, frowning, crying, moaning, and/or yelling)". Regarding the items of ELICSE, PosGaze_LN focuses on annotating positive facial expressions toward the stimulus, whereas PosReach_LN emphasizes positive affective touch, as to say manipulations of the artifacts in the activity (i.e., the robot and the interactive display) that have a positive affective nuance (e.g., stroke the robot).

The above-detailed descriptions could help understand why we found a significant difference on the item Atti_M but not on PosGaze_LN and *Pleasure*. The former could be explained by participants' significantly increased overall behavioral engagement towards the activity when the auditory feedback was added. The latter might indicate that this difference was not present when the single modality of facial expressions was taken into account. There are two other possible reasons besides the assumption that there was simply no difference in positive facial expressions between the two levels of multimodal stimuli. First, as PWD are often affected by impaired emotion regulation, some participants might have found it difficult to express their emotions through facial expressions. Further analyses could be performed with participants clustered per emotional disorders. Second, the sample size was too small for discovering statistical significance, under which circumstances more participants need to be recruited in future studies.

In addition, the results of Soc_E and Vis_E from the EPWDS also showed significant main effects of multimodal stimuli. According to the manual [37], the item Soc_E evaluates the interpersonal social interaction by measuring whether the participants used the activity provided as a communication channel to interact with others (as we have considered the HRI as part of the activity engagement). Hence, as the participants were more willing to verbally communicate with the facilitator when auditory stimuli were presented, social engagement with the facilitator increased as well. For Vis_E, it differs in that Gaze_LN only focuses on gaze behaviors directed towards the stimulus/activity, while visual engagement of the EPWDS also measures eye-contact with the person/s involved. The results could be explained by a consequence of the increased social activity with the facilitator. The discussion further confirmed that sensory enrichment has the potential in promoting not only activity-related engagement but also social engagement with human partners within our specific context.

6.1.2 Lack of Significance on Level of System Interactivity and Interaction Effects

The statistical analysis of data collection using ELICSE did not reveal any statistically significant main effects on the level of interactivity or interaction of system interactivity level and level of multimodal stimuli on engagement (except for the total duration of sessions, see Table 6). We speculated two reasons for possible explanations. The first reason considers the participants' diverse heterogeneity and design of experimental procedures. Specifically, how the provided activity should be presented to participants with different cognitive abilities so that they have a better understanding of all the functionalities and interaction possibilities of the system design. Dementia affects each participant differently. Our recruited participants were affected by behavioral disorders varying in severity and type. Participants with more advanced stages of dementia have a higher risk of not recognizing or increased difficulty in recognizing the increased system interactivity design due to narrower attention span and inability to notice the changes in the conditions, especially when only visual feedback were presented on the screen display (i.e., as in condition M1I2). Hence, the logical connection between interacting with the robotic sheep and the responsive feedback from another location - the screen display - could be difficult for participants with a high level of cognitive impairment to comprehend. In the implemented procedure design, we have arranged a brief introduction by the facilitator about how the designed system works pre-interaction session verbally. The intention was to retain the self-exploration, which aims to reinforce the rewarding experience when users successfully discovered the connection between touch input on the robot and feedback from the display themselves. How-

ever, in practice, such a connection might not be perceived by every participant, and this highly depends on their condition. Therefore, elaborate demonstrations by the facilitator and necessary guidance during the sessions could be useful for a better understanding of the logic connections, especially for participants with more advanced conditions.

The second reason for lack of significance regards the system implementation of the robotic animal design and facilitation of the HRI. The robotic sheep consisted of PLEO robot with sheep clothes and several touch sensors were embedded on the back, rear, head, and chin of the robot. During study sessions, not every touch input on the robot successfully triggered the programmed responses (e.g., when participants were petting the tail or legs). Hence, proper facilitation is crucial in guiding the participants through the designed feedback. Not enough exposure to responsive feedback could also be the reason for the lack of significance. As in this study, the robot was covered in sheep-like fur, and future studies could use textile embedded sensors for better coverage of the surface of the robot to ensure a more sensitive collection of user input. Furthermore, the facilitation of the HRI is also crucial in determining the positive effects. In some cases, we have noticed that certain participants seemed to fail in distinguishing whether the robot was on or off. In other words, unless been constantly addressed and guided by the facilitator on how the robot behaves and reacts, the users are at risk of not knowing the feedback from the robots or even not able to tell whether it is a robot or a real animal. As most traditional therapeutic interventions for PWD are often performed by specialists with professional training, the facilitation of robot use should also consider setting up standards for proper guidance and ethics to have its desired positive impact on dementia users.

Nevertheless, the non-significant results did not necessarily suggest there were no positive effects of increased interactivity of system design on user engagement. The results of ELICSE-based assessment showed a trend of increased positive gaze, and positive reach out behaviors (see Table 6), as well as more evident pleasure (see Table 7) when the system interactivity was higher. It is well known that the failure to demonstrate statistical significance may also be the result of low statistical power when an important effect actually exists and the null hypothesis of no effect is in fact false. However, due to the controversy of reporting the post hoc power calculation in literature (see the work of [30] for a complete discussion), we did not perform post hoc power calculations to aid the interpretation of non-significant results but reported a priori power calculation to guide the sample size instead (see sect. 4.2).

Taken together, this discussion provides more detailed insights on how multimodal stimuli presentations could influence the engagement of PWD under the specific contextual interaction design of this study. In its most direct sense,

increased experienced richness at a sensory level influences PWD's engagement by promoting manipulation of the social robot with positive emotions and facilitating communication with the human partner, which further leads to an increased attitude towards the activity and social engagement with the facilitator. In addition, our study showed that designing proper system interactivity requires careful considerations, as there is a need to balance the residual abilities of PWD with the amount of interactive possibilities that the system offers. To accommodate each user's unique conditions and allow users with dementia with different deterioration levels to benefit from the provided activity, it is essential that the activity is appropriately introduced and constantly facilitated throughout the whole session. In conclusion, the findings as mentioned above indicate that an increased sensory richness and richer interaction possibilities of an activity design can lead to a more positive attitude towards the activity, and could be used as motivation strategies for initiating and facilitating engagement, maintaining user interests, and facilitating verbal communications of PWD.

6.2 Discussion on Engagement Assessment Using ELICSE

The previous section regarding affective states assessment has already given a glimpse of how interpretation could vary due to the different measurement tools used, which primarily demonstrated the necessity of mixed measures. Here we further explain the reasons for adopting the video coding analysis using the ELICSE besides the golden standards rating scales. Like all video coding analysis methodologies, the use of the ELICSE could be very time-consuming. Therefore, the question raised is: what are the time cost trade-offs for assessing participants' engagement using ELICSE? There are several advantages in this case. First, unlike other validated measures used in this study, the ELICSE was built with the intention to be modified according to the specific context of the interaction, participants, and type of activities. Such adaption takes the nature of activities (i.e., passive or active activity), whether there were social interactions involved (e.g., alone, with partners, or in a group), and the conditions of the participants (e.g., mobility) into consideration [54]. The modification is then made to ensure the final coding scheme is meaningful within the specific context and comprehensive enough for capturing the engagement of PWD. The ELICSE identifies engagement-related behaviors under a specific context of interaction and helps the researcher to associate a meaning (i.e., *Attention* or *Valence*) to each specific behavior s/he observes. Second, the nature of coding analysis allows the quantification of each behavior more objectively and, hence, more robustly. Also, unlike post-experiment measures that require raters to recall previous experiences, the ELICSE is rated by taking the ongoing process of the interaction into

account, which reduces the risk of human error and wrong impressions. Last and most importantly, the ELICSE provides comprehensive details that allow researchers to further aggregate and conduct statistical analysis according to specific research goals. For instance, in this study, we aim to investigate how the user engagement with the provided activity could be influenced by different system configurations applied in designed experimental conditions, regardless of their social interaction with the facilitator. Therefore, data aggregations of the ELICSE were performed according to their attention focus on the activity only.

Next, we discuss the reasons for presenting the results using both absolute duration and percentage duration (i.e., calculated using absolute duration/ length of the total duration of a session) of the behaviors in the ELICSE, and possible underlining reasons for different results. The results presented in Table 6 showed a significant main effect of multimodal stimuli level on PosReach_LN using absolute duration data collection. However, no significance was found using the percentage duration of the same item. Similarly, the significant main effect of multimodal stimuli level on item TalkAct were only exhibited when using percentage duration. The different results suggest that the two ways of data collection measure engagement differently.

The percentage data collection calculates the proportion of a particular behavior/modifier out of the whole session. It has the advantage of even out the influence of a session's total duration by computing a percentage that demonstrates a direct impression on the user's focus distributions. However, in practice, when participants are less interested in the provided activity, they naturally shift their attention towards the facilitator for interpersonal interaction. The interaction with the facilitator would influence the final results using percentage duration. More specifically, participants may gaze towards the facilitator more if they had recollected memories triggered by the interaction and wished to share their experience with the facilitator, consequently reducing the percentage of gaze towards the screen or robotic sheep. To address this, we have also exported analyzed data using absolute duration. The duration of time that participant was occupied or involved with a stimulus, suggested by Cohen-Mansfield et al. [14,16], is an essential indicator of user engagement of PWD. Absolute duration data takes the total duration length of an interaction session into consideration, and aims to reflect the extent to which the participant is willingly spending their time with the stimulus regardless of the rapport behaviors with the facilitator. In this sense, the results of bodily behaviors using absolute duration are closer to reflect the nature of activity-related engagement.

Regarding verbal behaviors - *Conversations* of ELICSE, they are different in nature from bodily expressions. Most of the verbal expressions occurred between two human partners (i.e., the facilitator and the participant), except self-mumbling

or talking to the sheep (both as a robot or as screen content). Hence, we could not separate the facilitator's potential influence when performing data analysis but aggregated it into an item of TalkAct. We then analyzed the percentage of the total verbal expression for a general impression of communications.

6.3 Implications for Human-Robot Interaction Research within Dementia Care

Given the trend of global population aging, inflated healthcare costs, and lack of resources in most LTC facilities, there is a large likelihood that older adults with dementia will be accompanied by robots in the future, whether for assisting independent living or fulfilling psycho-social needs. In this section, we present the implications derived from our findings that might be inspiring for future HRI research within the dementia care field.

Multisensory experience design for HRI. Most recent design research for PWD are sensory-based in their essence [32,65]. Social robots engage PWD in sophisticated multisensory ways to increase activity levels both from physical and social perspectives. On the one hand, recent robotics research is looking for a way to design the HRI experience so that it is more sensory holistic and immersive [1], on the other hand, studies have started to pay attention to how robot use could help to shape the everyday living experiences for elderly [24]. The presented study was conducted based on a specific activity design of context-enhanced HRI. We employed a responsive display to provide contextual information and sensory cues for a more immersive and richer HRI experience. In this way, the system design could benefit users not only from a sensory-stimulating way but also by creating a story narrative and a use context for robot facilitation and acceptance. Although the study has not investigated to which extent adding the artificial context on HRI contributes to the significant main effect of multimodal stimuli on enhancing engagement, it could perhaps offer a new perspective on HRI experience design by enabling multimodal feedback from a larger scale setting than the robot itself.

Adaptive system design with multiple interaction possibilities. Our activity design provides multiple interaction possibilities ranging from a simple "outlook experience" at the media content displayed on the screen, to "social robot petting" with HRI, and an "immersive sensory experience" that involve both robot interaction and interactive media content. These adapted levels of interaction allow users to freely explore the system design without the concerns of making mistakes, and compose their interaction in the way they are more comfortable with. The multiple interactive possibilities have the potential to adapt to various user conditions regarding various cognitive abilities but personal characteristics (e.g., mood during interaction). For instance, when users

are in agitated conditions, the "outlook experience" could provide relaxation and enjoyment. When they are bored and searching for stimulation, the interactive system could provide a social agent that acts as a companion and simulates human-animal therapeutic interaction. Moreover, the interactive system design could also help maintain the user interest and attentiveness, as users can continuously shift their attention between the dynamic media content shown on display and the robot behaviors to remain in flow. In addition, it could also help lower the barrier for physical and cognitive requirements since users in their wheelchair could also benefit from the low threshold physical interaction of cuddling and petting the robotic sheep.

Crucial role of facilitation during robot interaction. Like most occupational therapies developed for PWD to improve quality of life, interaction with a robot should also value the facilitation of specialists. First, the quality and conditions of facilitation are known to influence user's acceptance, and attitudes towards the robot interaction experience [27]. Second, there is a rich body of work that addresses the ethical concerns of social robots' use in dementia care, as it may tend to replace human care and lead to reduced human contact. The facilitation of a human caregiver is crucial as it could help maintain human contact within the HRI experience while lowering the risk of caretaking stress. It requires less focused attention and helps maintain the human-human interaction channel open. The support of a facilitator could also prevent other ethical concerns such as deception (recognize the robot as a real animal) [69], or infantilization feeling (similar to an adult who plays with a toy). In our research, we view the robot interaction not only as a stimulus for keeping PWD stimulated and improving their mood but also as a meditating artefact for interaction between humans-and-humans. For designers and developers of robotic research, we would suggest to consider making guidelines for human facilitation, and to do so by (1) carefully consider the dynamic relationship among human facilitator, provided stimuli, and users with dementia, and how it would shape the daily living for multi-stakeholders; also, (2) as recent research starts to involve dementia users in inclusive design processes [72], we would suggest to include caregivers/facilitators in the initial developing process as well.

6.4 Limitations and Future Work

The major limitation of this study lies in the relatively small sample size and uneven participants' distribution in the two groups. The small sample size was due to challenges in the recruitment of PWD in the relatively small community of Vitalis, and the uneven group size was due to the participants' withdrawal during the study. The sample size was also a result of considerations influenced by choice of methodologies, which requires a significant amount of effort and

investment in time. Given the above practical limitations, partial eta-squared effect size values were reported to substantiate the scope of the results. Future work should attempt to replicate the experiment with a larger sample size and participants from different locations of nursing homes. Furthermore, due to the small sample size and the low number of participants for each level of dementia severity, we could not perform further statistical analyses focusing on the effects of participants' characteristics on engagement. Future work should consider recruiting a larger number of participants in each level of dementia severity to examine differences caused by the disease's progression. Additionally, since user's facial expressions could also be hindered by emotional disorders, future work should analyze the effect of users' affective disorders (e.g., depression, apathy, anxiety) on their facial expressivity to make the assessment of engagement more sound.

Moreover, there is an uneven sample size between genders (12 female participants and 4 male participants), which potentially giving the impression of gender bias when interpreting the results. In fact, the majority residents living in Vitalis are women and so are many other nursing homes worldwide (see reference for instance [8]). According to the literature, there exists a gender difference among population of residents with dementia living in nursing homes. The admission rates between the male and female ratio ranged between 1 to 1.4 and 1 to 1.6 according to international studies reported by Luppa et al. [44]. And the gender difference in nursing home placements of PWD is generally explained by the higher life expectancy of women at present, the slightly higher dementia prevalence rates of women than man (10.1% vs. 9.6%) [23], the higher rate of women living alone in older age than man [44], and the tendency of willingly to give care of women than man. Overall, we believe that our sample of this study represents the gender profile of nursing home residents with dementia disease. And this raises the awareness of designing for gender differences, in particular, for older women with dementia in future works.

Other limitations concern the measures and data analysis. First, due to practical consideration, the rating scales were filled out based on varied materials (OME, OERS were rated based on direct observations on-site, while EPWDS were rated based on video recordings) by two different raters. This might have slightly weakened consistency among the three rating scales. However, it did not influence the reliability of results as the use of a single scale was consistent across all experimental sessions. Second, the video coding analysis using the ELICSE adopted both percentage and absolute duration due to the distinct length of each session and mutual dependence between activity-related engagement and interpersonal interaction with the facilitator. Future studies should make clear guidelines of experiment design and procedures to ensure the robustness of ELICSE video coding analy-

sis using percentage data representation. Instructions such as try to reduce personal conversations that are irrelevant to the study with the participants and try to be consistent and follow the same study procedure for all participating sessions could be implemented. Lastly, although the mixed method used in this study yielded a reliable assessment of user engagement, future work should consider combining this mixed method with a more qualitative interpretation of participants' behaviors by people who entertain trustful relationships with residents. The meaning of the annotated behaviors would be increased by the understanding of each participant from a person-centered care point of view [71]. Future work could collect participants' lifestyle, personality, preference, and past/present interests using tools such as the Self-Identity Questionnaire proposed by Cohen-Mansfield et al. [15,17] for a better interpretation of the user engagement.

Furthermore, as this study invited one participant at a time to better control the experimental conditions, future work should also test how the system adapts to a pair of users and how their activity-related engagement and human-human interaction would be facilitated by context-enhanced HRI. The system design presented in this paper adopted visual-auditory-tactile feedback for multimodal stimuli presentation. Future work could attempt to engage more sensory channels (e.g., aroma-diffuser of grass field for olfactory display) of PWD for a more holistic and realistic sensory experience. Additionally, the robotic sheep design in this study could be further improved by adding heating elements. As suggested by Block et al. [6] that physical warmth helps promote social warmth, the adoption of heating features besides inviting texture and appealing appearance are likely to promote HRI.

7 Conclusion

To address the current disengaged and under-stimulated living situation of PWD in LTC facilities, this study attempted to explore how to design rich interaction experience to improve the level of engagement of PWD. The experiment design was built on a previous work of *LiveNature* design that echos to the nostalgic experience of a generation of Dutch elderly and utilized intuitive interfaces that users already familiar with. The system design suggested a novel approach - context-enhanced HRI - which combined the interaction with a tangible social robot with an augmented reality responsive display. With social robots being increasingly employed in the complex domain of dementia care, this study investigated the role of multimodal stimuli and interactivity in improving the richness of the experience based on the approach of context-enhanced HRI. The sensorial level of experienced richness was addressed by the system design's multimodality sensory feedback. And the system interactivity was varied

based on whether the HRI was accompanied by contextual cues from the augmented reality display. The engagement of participants was assessed using a mixed assessment method involving the use of video analysis (using the ELICSE) and three observational rating scales (OME, OERS, and EPWDS). Results provide sufficient evidence of the significant contributing role of multimodal stimuli on improving emotional aspects of activity-related engagement and social interaction with a human partner. The findings could be potentially used as motivation strategies in future design research for promoting PWD's positive attitude, facilitating communications, and social rapport. It could also contribute to several domains of knowledge, namely: (1) the domain of interaction design for dementia. While most sensory-based designs for PWD were mainly focused on the stimulation of certain senses, for instance, music/sound for reminiscence or textile designs for comforting and relaxation. This research addresses the significant benefits of employing multimodal sensory presentations, including dynamic visual content, auditory stimulation, and tactile explorations; In addition, (2) it contributes to robotic research by offering a novel way of combining sensory cues embedded in environmental settings with the HRI, and addressing the critical role of professional facilitation; lastly, (3) it adds insights to dementia engagement study by providing a comprehensive mixed method for engagement assessment.

Declarations

References

1. Alenljung B, Lindblom J, Andreasson R, Ziemke T (2019) User experience in social human-robot interaction. In: Rapid automation: Concepts, methodologies, tools, and applications, IGI Global pp 1468–1490

2. Anderiesen H, Scherder E, Goossens R, Visch V, Eggermont L (2015) Play experiences for people with alzheimers disease. International Journal of Design 9(2):2015

3. Bainbridge WA, Hart JW, Kim ES, Scassellati B (2011) The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot 3(1):41–52

4. Bemelmans R, Gelderblom GJ, Jonker P, De Witte L (2012) Socially assistive robots in elderly care: A systematic review into effects and effectiveness. J Am Med Dir Assoc 13(2):114–120

5. Bennett B, McDonald F, Beattie E, Carney T, Freckelton I, White B, Willmott L (2017) Assistive technologies for people with dementia: ethical considerations. Bull World Health Organ 95(11):749

6. Block AE, Kuchenbecker KJ (2019) Softness, warmth, and responsiveness improve robot hugs. Int J Soc Robot 11(1):49–64

7. Broekens J, Heerink M, Rosendal H et al (2009) Assistive social robots in elderly care: a review. Gerontechnology 8(2):94–103

8. Buchanan RJ, Wang S, Ju H, Graber D (2004) Analyses of gender differences in profiles of nursing home residents with alzheimers disease. Gender Medicine 1(1):48–59

9. Campo M, Chaudhury H (2012) Informal social interaction among residents with dementia in special care units: Exploring the role of the physical and social environments. Dementia 11(3):401–423

10. Chen SC, Jones C, Moyle W (2018) Social robots for depression in older adults: a systematic review. J Nurs Scholarsh 50(6):612–622

11. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Measur 20(1):37–46

12. Cohen J (2013) Statistical power analysis for the behavioral sciences. Academic press, Cambridge

13. Cohen-Mansfield J (2013) Nonpharmacologic treatment of behavioral disorders in dementia. Curr Treat Options Neurol 15(6):765–785

14. Cohen-Mansfield J, Dakheel-Ali M, Marx MS (2009) Engagement in persons with dementia: the concept and its measurement. Am J Geriatr Psychiatry 17(4):299–307

15. Cohen-Mansfield J, Golander H, Arnheim G (2000) Self-identity in older persons suffering from dementia: Preliminary results. Soc Sci Med 51(3):381–394

16. Cohen-Mansfield J, Marx MS, Freedman LS, Murad H, Regier NG, Thein K, Dakheel-Ali M (2011) The comprehensive process model of engagement. Am J Geriatr Psychiatry 19(10):859–870

17. Cohen-Mansfield J, Marx MS, Thein K, Dakheel-Ali M (2010) The impact of past and present preferences on stimulus engagement in nursing home residents with dementia. Aging Ment Health 14(1):67–73

18. Cohen-Mansfield J, Perach R (2015) Interventions for alleviating loneliness among older persons: a critical review. Am J Health Promot 29(3):e109–e125

19. Erdfelder E, Faul F, Buchner A (1996) Gpower: A general power analysis program. Behavior Res Methods, Instrum, Comput 28(1):1–11

20. Feng Y, Barakova EI, Yu S, Hu J, Rauterberg G (2020) Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. Sensors 20(13):3771

21. Feng Y, Yu S, van de Mortel D, Barakova E, Hu J, Rauterberg M (2019) Livenature: ambient display and social robot-facilitated multi-sensory engagement for people with dementia. pp 1321–1333

22. Fleiss JL, Levin B, Paik MC (2013) Statistical methods for rates and proportions. Wiley, NY

23. Freedman VA, Kasper JD, Spillman BC, Plassman BL (2018) Short-term changes in the prevalence of probable dementia: An analysis of the 2011–2015 national health and aging trends study. J Gerontol: Ser B 73(1):S48–S56

24. Frennert S, Östlund B (2014) Seven matters of concern of social robots and older people. Int J Soc Robot 6(2):299–310

25. Góngora Alonso S, Hamrioui S, de la Torre Díez I, Motta Cruz E, López-Coronado M, Franco M (2019) Social robots for people with aging and dementia: a systematic review of literature. Telemed e-Health 25(7):533–540

26. Hancock GA, Woods B, Challis D, Orrell M (2006) The needs of older people with dementia in residential care. Int J Geriatric Psychiatry: A J Psychiatry Late Life Allied Sci 21(1):43–49

27. Heerink M, Kröse B, Evers V, Wielinga B (2010) Assessing acceptance of assistive social agent technology by older adults: the almere model. Int J Soc Robot 2(4):361–375

28. Hendrix J, Feng Y, van Otterdijk M, Barakova E (2019) Adding a context: will it influence human-robot interaction of people living with dementia? Springer, pp 494–504

29. Higashino K, Kimoto M, Iio T, Shimohara K, Shiomi M (2020) Effects of social touch from an agent in virtual space: Comparing visual stimuli and virtual-tactile stimuli. pp 768–774

30. Hoenig JM, Heisey DM (2001) The abuse of power: the pervasive fallacy of power calculations for data analysis. Am Stat 55(1):19–24

31. Houben M, Brankaert R, Bakker S, Kenning G, Bongers I, Eggen B (2020) The role of everyday sounds in advanced dementia care. pp 1–14

32. Houben M, Brankaert R, Kenning G, Eggen B, Bongers I (2020) The perspectives of professional caregivers on implementing audio-based technology in residential dementia care. Int J Environ Res Public Health 17(17):6333

33. Howell AJ, Dopko RL, Passmore HA, Buro K (2011) Nature connectedness: Associations with well-being and mindfulness. Personality Individ Differ 51(2):166–171

34. Huber S, Berner R, Uhlig M, Klein P, Hurtienne J (2019) Tangible objects for reminiscing in dementia care. pp 15–24

35. Hung L, Liu C, Woldum E, Au-Yeung A, Berndt A, Wallsworth C, Horne N, Gregorio M, Mann J, Chaudhury H (2019) The benefits of and barriers to using a social robot paro in care settings: a scoping review. BMC Geriatr 19(1):232

36. Jones C, Sung B, Moyle W (2015) Assessing engagement in people with dementia: a new approach to assessment using video analysis. Arch Psychiatr Nurs 29(6):377–382

37. Jones C, Sung B, Moyle W (2018) Engagement of a person with dementia scale: Establishing content validity and psychometric properties. J Adv Nurs 74(9):2227–2240

38. Kaplan R (2001) The nature of the view from home: Psychological benefits. Environ Behav 33(4):507–542

39. Laban R (2011) Choreutics. alton. Hampshire: Dance Books Ltd.(Original work published 1966)

40. Lawton MP, Van Haitsma K, Klapper J (1996) Observed affect in nursing home residents with alzheimer's disease. J Gerontol Ser B: Psychol Sci Soc Sci 51(1):P3–P14

41. Lazar A, Edasis C, Piper AM (2017) Supporting people with dementia in digital social sharing. pp 2149–2162

42. Lazar A, Thompson HJ, Piper AM, Demiris G (2016) Rethinking the design of robotic pets for older adults. pp 1034–1046

43. Livingston G, Kelly L, Lewis-Holmes E, Baio G, Morris S, Patel N, Omar RZ, Katona C, Cooper C (2014) Non-pharmacological interventions for agitation in dementia: systematic review of randomised controlled trials. Br J Psychiatry 205(6):436–442

44. Luppa M, Luck T, Weyerer S, König HH, Riedel-Heller SG (2009) Gender differences in predictors of nursing home placement in the elderly: a systematic review. Int Psychogeriatr 21(6):1015

45. McLaren AN, LaMantia MA, Callahan CM (2013) Systematic review of non-pharmacologic interventions to delay functional decline in community-dwelling patients with dementia. Aging Mental Health 17(6):655–666

46. Miranda-Castillo C, Woods B, Galboda K, Oomman S, Olojugba C, Orrell M (2010) Unmet needs, quality of life and support networks of people with dementia living at home. Health Qual Life Outcomes 8(1):1–14

47. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G (2013) Use of social commitment robots in the care of elderly people with dementia: A literature review. Maturitas 74(1):14–20

48. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ER, Shum DK, O'Dwyer ST, Mervin MC, Draper BM (2017) Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Direct Assoc 18(9):766–773

49. Neal I, du Toit SH, Lovarini M (2020) The use of technology to promote meaningful engagement for adults with dementia in residential aged care: a scoping review. Int Psychogeriatr 32(8):913–935

50. O'Neil ME, Freeman M, Christensen V, Telerant R, Addleman A, Kansagara D et al. (2011) A systematic evidence review of non-pharmacological interventions for behavioral symptoms of dementia. Department of Veterans Affairs Washington, DC

51. Organization WH (2012) Dementia: a public health priority. World Health Organization, Geneva

52. Perugia G, Díaz-Boladeras M, Català-Mallofré A, Barakova EI, Rauterberg M (2020) Engage-dem: a model of engagement of people with dementia. IEEE Transactions on Affective Computing

53. Perugia G, Doladeras MD, Mallofré AC, Rauterberg M, Barakova E (2017) Modelling engagement in dementia through behaviour. contribution for socially interactive robotics. IEEE, pp 1112–1117

54. Perugia G, Paetzel-Prüsmann M, Alanenpää M, Castellano G (2021) I can see it in your eyes: Gaze as an implicit cue of uncanniness and task performance in repeated interactions with robots. Frontiers in Robotics and AI 8

55. Perugia G, Van Berkel R, Díaz-Boladeras M, Català-Mallofré A, Rauterberg M, Barakova E (2018) Understanding engagement in dementia through behavior. the ethographic and laban-inspired coding system of engagement (elicse) and the evidence-based model of engagement-related behavior (emodeb). Front Psychol 9:690

56. Rozendaal M, Keyson DV, de Ridder H (2007) Interactivity, vividness and richness in user interfaces. Trento, Italy, pp 269–278

57. Rozendaal MC (2007) Designing engaging interactions with digital products. Ph.D. thesis, Mechanical Maritime and Materials. Engineering

58. Rozendaal MC, Keyson DV, de Ridder H, Craig PO (2009) Game feature and expertise effects on experienced richness, control and engagement in game play. AI Soc 24(2):123–133

59. Rozendaal MM, Keyson DV, de Ridder H (2009) Product features and task effects on experienced richness, control and engagement in voicemail browsing. Pers Ubiquit Comput 13(5):343–354

60. Šabanović S (2010) Robots in society, society in robots. Int J Soc Robot 2(4):439–450

61. Samus QM, Rosenblatt A, Steele C, Baker A, Harper M, Brandt J, Mayer L, Rabins PV, Lyketsos CG (2005) The association of neuropsychiatric symptoms and environment with quality of life in assisted living residents with dementia. Gerontologist 45(1):19–26

62. Schölzel-Dorenbos CJ, Meeuwsen EJ, Olde Rikkert MG (2010) Integrating unmet needs into dementia health-related quality of life research and care: Introduction of the hierarchy model of needs in dementia. Aging Ment Health 14(1):113–119

63. Steuer J (1992) Defining virtual reality: Dimensions determining telepresence. J Commun 42(4):73–93

64. Strøm BS, Ytrehus S, Grov EK (2016) Sensory stimulation for persons with dementia: a review of the literature. J Clin Nurs 25(13–14):1805–1834

65. Thoolen M, Brankaert R, Lu Y (2020) Ambientecho: exploring interactive media experiences in the context of residential dementia care. pp 1495–1508

66. Topo P (2009) Technology studies to meet the needs of people with dementia and their caregivers: a literature review. J Appl Gerontol 28(1):5–37

67. Trahan MA, Kuo J, Carlson MC, Gitlin LN (2014) A systematic review of strategies to foster activity engagement in persons with dementia. Health Edu Behav 41(1):70S-83S

68. Treadaway C, Taylor A, Fennell J (2019) Compassionate design for dementia care. Int J Design Creat Innov 7(3):144–157

69. Van Maris A, Zook N, Caleb-Solly P, Studley M, Winfield A, Dogramadzi S (2020) Designing ethical social robots–a longitudinal field study with older adults. Frontiers in Robotics and AI 7

70. Wada K, Shibata T, Musha T, Kimura S (2008) Robot therapy for elders affected by dementia. IEEE Eng Med Biol Mag 27(4):53–60

71. Wallace J, Wright PC, McCarthy J, Green DP, Thomas J, Olivier P (2013) A design-led inquiry into personhood in dementia. pp 2617–2626

72. Wang G, Marradi C, Albayrak A, van der Cammen TJ (2019) Co-designing with people with dementia: A scoping review of involving people with dementia in design research. Maturitas 127:55–63

73. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D (2018) Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J Soc Robot 10(5):643–668

74. Yamaguchi H, Maki Y, Yamagami T (2010) Overview of non-pharmacological intervention for dementia and principles of brain-activating rehabilitation. Psychogeriatrics 10(4):206–213
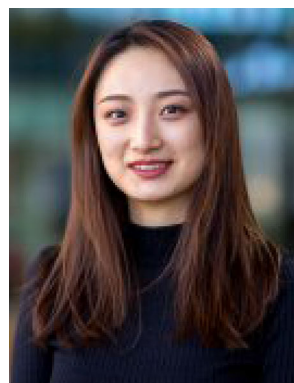
Yuan Feng is currently a double PhD candidate in the department of Industrial Design at Eindhoven University of Technology (TU/e, the Netherlands) and Northwestern Polytechnical University (NPU, China). She received both her BEng and MSc degrees in industrial design from Northwestern Polytechnical University in 2012 and 2015 respectively. Her recent research interests include multimodal interaction, human-robot interaction, and design for health and well-being of elderly users including people with dementia. She is a member of the Social Robotics Lab Research Group of industrial design TU/e and TU/e Design for Social Innovation and Sustainability (DESIS) Lab.

Giulia Perugia (she/her) is an Assistant Professor (tenure track) in Human-Robot Interaction at the Human-Technology Interaction Group of Eindhoven University of Technology (TU/e, Netherlands). Previously, she was a postdoctoral researcher at Uppsala University (Sweden) and an Erasmus Mundus Ph.D. student at the Polytechnic University of Catalonia (UPC, Spain) and TU/e. She holds a BA in Literature and Linguistics (University of RomaTre, Italy), a M.Sc. in Cognitive Science (University of Siena, Italy), and a double degree Ph.D. in Assistive Technologies (UPC and TU/e). Her research lies at the intersection of Human-Robot Interaction (HRI), Affective and Multimodal Interaction, and Educational and Assistive Technologies. She was awarded the Distinguished Interdisciplinary Research Award by the Robotic Society of Japan and the Korean Robotics Society in 2017 and was shortlisted for the same award in 2020.

Suihuai Yu is a professor in the department of Industrial Design at Northwestern Polytechnical University (NPU, China). As a doctoral supervisor, Prof. dr. Yu is currently the Vice President of China Industrial Design Association, the Director of Industrial Design and Ergonomics Key Laboratory under Ministry of Industry and Information Technology of China, the Deputy Director of Teaching and Instruction Subcommittee of Industrial Design Department under Ministry of Education of China, and the Director of Shaanxi Industrial Design Engineering Laboratory.

Emilia I. Barakova received a Ph.D. degree in mathematics and physics from Groningen University, Groningen, The Netherlands, in 1999 and holds a Masters' degree from Technical University Sofia, Bulgaria. She is currently with the Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands, and leads the Social Robotics Lab of the Eindhoven University of Technology. She has hold research positions at RIKEN Brain Science Institute (Japan), GMD-Japan Research Laboratory (Japan), Groningen University (The Netherlands), and the Bulgarian Academy of Science (Bulgaria). She is an associate editor of International Journal of Social Robotics and Editor of Personal and Ubiquitous Computing and has organized several IEEE and ACM conferences. She has expertise in social robotics, AI and robotics, modeling social behavior, and human-centered interaction design.
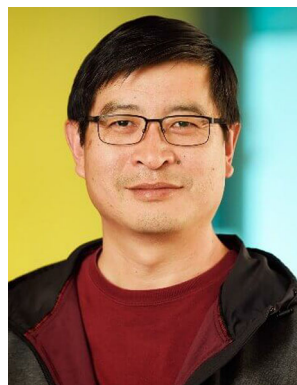
Jun Hu is a Senior Member of ACM, currently an Associate Professor in Design Research on Social Computing, and the Scientific Director for the Professional Doctorate in Engineering program in User System Interaction at Eindhoven University of Technology (TU/e), a Distinguished Adjunct Professor at Jiangnan University and a Guest Professor at Zhejiang University. He has a Ph.D. in Interaction Design and a Professional Doctorate in User-system Interaction, both from TU/e. He also has a B.Sc in Mathematics and an M.Eng in Computer Science. His research interests are in HCI, IoT, AI, industrial design, computer science, and design education.

Matthias Rauterberg received the B.S. degree in psychology from the University of Marburg, Marburg, Germany, in 1978; the B.A. degree in philosophy in 1981, the B.S. degree in computer science in 1983, the M.S. degree in psychology in 1981, and the M.S. degree in computer science in 1986 from the University of Hamburg, Hamburg, Germany; and the Ph.D. degree in computer science/mathematics from the University of Zurich, Zurich, Switzerland, in 1995. He was Professor and Head of the Designed Intelligence Research Group, Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands. His research interests include entertainment computing, cultural computing, human-computer interaction, and design science. He is a fellow of the International Federation for Information Processing (IFIP).