Robot Assisted Sing-along for Groups of Individuals with Dementia

Real-time Engagement Detection and Re-engagement in Human Robot Interaction

Manali Shah

Robot Assisted Sing-along for Groups of Individuals with Dementia

Real-time Engagement Detection and Re-engagement in Human Robot Interaction

by

Manali Shah

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday June 25, 2024 at 10:00 AM.

Student number:	5739896	
Project duration:	September 5, 2023 – June 25, 2024	
Thesis committee:	Prof. Dr. Mark Neerincx	TU Delft, supervisor
	Dr. Bernd Dudzik	TU Delft, supervisor
	Dr. Chirag Raman	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Abstract

Cognitive Impairment, commonly termed as Dementia affects a large number of older adults. People with dementia (PwD) experience cognitive decline that impacts their ability to perform daily activities and maintain social connections. The number of PwD is expected to rise, and unfortunately, there is still no cure. A number of non drug therapies such as music and art therapy are used to stimulate neurons and slow down the progress of the disease. When symptoms become severe, full-time care may be needed, requiring a move to a nursing home for formal care. However, studies show that interaction with peers is limited in a care home, leading to loneliness and reduced quality of life. Caregivers organize group activities like singing and baking to alleviate these feelings, but their busy schedule can limit the frequency of these exercises. This begs the question: Can a robot support caregivers in the provision of group activities in a care home? This would allow caregivers more time to attend to the personal needs of the residents. It not only has the potential to reduce workload of caregivers but also creates opportunities for residents to interact with each other.

In this thesis, we propose the design of a robot which moderates a sing-along activity for a group of people living in a nursing home. An activity in the realm of music therapy is chosen since it is one of the non-drug therapies which has shown to have numerous benefits like improving memory recall, eliciting emotions and aiding relaxation. The robot session is curated to give opportunities for interaction with group members and with the robot too. While the song is playing and participants are singing along with the music, the robot performs engagement detection, and encourages those who do not seem involved.

Multiple ways are proposed to detect engagement in HRI, such as emotion recognition and estimation of head pose. These approaches have limitations in the scope of this thesis, particularly because these methods are not designed for groups of older adults residing in a care home. Given these limitations and our focus on the target group, we propose a novel engagement detection methodology, which leverages the presence of a robot and assigns an active role to it. Instead of relying on conventional passive engagement detection methods, this approach uses an interactive probing technique. The robot prompts users to perform certain actions, and whether or not they respond to it is monitored using pose estimation. This is termed robot engagement. This approach is combined with activity engagement, which indicates whether the participant is singing or not. With this hybrid technique, we aim to increase the efficacy of engagement detection. If a participant is detected as disengaged, either in the activity or in the robot, we provide personalized encouragement and motivation by addressing participants by name. Along with this, the robot also compliments those who actively participate, motivating them further.

Since this is a first exploratory study, evaluation is not performed with the original (vulnerable) target audience. Instead, the prototype is tested with older adults over the age of 60. With this, we gather feedback and model design recommendations for future iterations of the system. Along with this, we gain insights into the feasibility of the proposed engagement detection method. Feedback about the session is collected through three data sources: a subjective questionnaire, an objective questionnaire and a behavioural analysis of the video recordings. From this data, we can make several observations. It is seen that while participants are initially intrigued about the robot, they do not respond to the robot's prompts. We also observe that personalized comments are important in maintaining engagement, along with a robot with more human-like features. Additionally, the pose estimation algorithm proposed shows promise, but requires testing to make it robust. A high level of activity engagement is noticed, with participants singing all of the songs in their entirety, which could be attributed to a selection bias during recruitment. Finally, we note that the robot provides a platform for them to interact with each other, thereby enhancing social engagement.

Noting the limitations of the experiments and reflecting upon the feedback received, recommendations are made to improve the system in future iterations. Ensuring the robot's speech is audible over the music, while also making it more organic through the use of LLMs, is imperative for future systems. Important considerations while using the pose estimation algorithm are discussed. Additionally, suggestions are made to enhance personalization in the interaction, which is vital in HRI. Finally, the robot's behaviour and appearance must be humanized for increased acceptance into human teams.

Acknowledgements

As my thesis suggests, humans are social beings who depend on each other for support and guidance. Likewise for me, there were multiple individuals throughout the course of my thesis whose support was essential to its completion.

- Firstly, I would like to thank my parents and my brother, Shubh, who have worked tirelessly to provide with me the opportunity to study abroad at a prestigious university like TU Delft. Leaving my family and moving across continents was not an easy decision, but their unwavering support and faith in me inspires me to work harder to make them proud. My father, born with a hearing impairment, kindled the desire to pursue research in a field for individuals with special needs, and I am happy to have realized that ambition through this thesis.

- I had the privilege to work under the supervision of not two, but three supervisors during my thesis: Prof. Dr. Mark Neerincx, Dr. Frank Broz and Dr. Bernd Dudzik. Performing a detailed research project was something I had never done before but they patiently guided me through the smallest of doubts. During my weekly meetings, they provided insightful feedback which has been invaluable to improve my research. Their attention to detail and commitment for perfection made me strive to achieve the highest standards in my work, leading to the successful completion of this thesis.

- A big thank you to Frits van Hall, the student assistant who helped me recruit participants for the experiment by making posters, distributing it to choirs in Delft and coordinating the schedule of interested participants. Not only did he translate the all the forms and questionnaires to Dutch, but also assisted during the experiments. When participants arrived for the experiments, he made them comfortable and helped them settle down, while I focused on setting up the experiment.

- My partner, Abhiroop, has been a persistent source of emotional support. Apart from proof-reading my thesis and listening to my mock presentation numerous times, he has been the one who motivated me to keep going whenever my enthusiasm wore off. He kept me centered and focused, urging me to do my best every single day.

- During the course of my master's program, I worked on projects with the 'Dream Team', 4 remarkable girls who got me through my coursework. Each brought their unique expertise, providing learning experiences which I carried through my thesis. They made me a better researcher and enhanced my academic journey in so many ways. During the thesis, I was lucky to work with two of them, who also worked in HRI: Anagha and Shambhawi. Spending countless hours in the lab with them not only involved an exchange of ideas and feedback, but gave me the reassurance that I wasn't alone in this project.

- Finally, a big thank you to all my friends, those attending my defense and everyone online. Thank you for the game nights, movie nights and McDonald nights which provided much needed stress relief amidst our busy schedules!

List of Acronyms

SAR Socially Assistive Robots

HRI Human-Robot Interaction

PwD People with Dementia

RQ Research Question

WoZ Wizard-of-Oz

FER Facial Expression Recognition

QoL Quality of Life

RoSAS Robotic Social Attributes Scale

GUI Graphical User Interface

SIC Social Interaction Cloud

LLM Large Language Model

Contents

1 Introduction

The emergence and integration of robots into human lives has become notable in recent years. Not only are they used in factories and industries to perform challenging and intricate tasks, but have also become widespread in day-to-day activities. Robots can take various forms: medical robots, autonomous vehicles, vacuum cleaners and educational robots are just to name a few. They are intertwined with our lives in the form of smart home assistants like Alexa and Google Home [1]. Robots are being used in the medical field to perform surgeries and high precision tasks [2] as well as in the field of education as tutors, peers and tools [3]. Since the COVID-19 pandemic, it is observed that the prevalence of robots as social companions is on the rise [4]. As social isolation increased during the pandemic, robots were employed to reduce loneliness and provide companionship. This category of robots is called Socially Assistive Robots (SAR) [5].

1.1. Socially Assistive Robots

The class of robots designed to assist and interact with humans in a socially meaningful way is called Socially Assistive Robots (SAR) [5]. Instead of merely performing tasks, their main goal is to connect with humans on an emotional level through interactions [6]. They can take various forms, such as a humanoid robot, a service robot or a pet-like companion [7], depending on the situation and requirements. While over-humanizing a robot can result in negative implications, such as unrealistic expectations and feelings of existential threat, there are numerous benefits as well [8]. These include enhanced social interactions and well-being along with higher perceived support. Owing to these benefits, we work with the humanoid form of SARs in this thesis. These are robots with human characteristics that understand emotion, communicate with dialogue, build social relationships, use natural cues, have a personality, and learn social competencies [9]. They are also given a human like embodiment to enhance their lifelike attributes. Various robot sensors such as cameras and microphones along with artificial intelligence are utilized to enable SARs to understand and emulate human emotions. Gestures, body language, facial expressions and speech are to communicate with the users in meaningful and human-like ways.

SARs have found their calling in diverse fields, like healthcare, education and therapy. They cater to a wide variety of users including the elderly, individuals with physical impairments, persons with cognitive disorders, students and many more [6]. Providing human-like perception, emotion and dialog, they aim to simulate human companionship, thus enhancing overall mental health and well-being of the users. They are being used for children with autism to assist them with education and therapy [10], and for the elderly to help with loneliness and supervise those with mobility issues [11].

In this research, we explore Human Robot Interaction (HRI), SARs specifically, for persons with a diagnosis of dementia living in a care home. Interaction with the robot aims to serve two purposes for the target audience: improving cognitive functions with music therapy and reducing loneliness by facilitating interactions with fellow residents.

1.2. Cognitive Impairment

Cognitive impairment, more commonly termed as dementia, is a condition which causes damage to brain cells, thereby affecting functions of the brain [12]. This causes memory loss, influences decision making and causes a decline in thinking ability. The decline in cognitive ability can create a distance between the individuals and their family members, causing a sense of loneliness in them [13]. They are not able to communicate well, leading to limited social connections. The condition impacts a huge number of older adults, with the number of People with Dementia (PwD) expected to rise significantly in the near future.

Dementia is a progressive condition which can be classified into three stages: Early, Middle and Late stage dementia. At present, there is no cure for the condition; only the symptoms can be managed and the pace of decline can be slowed down. Thus, as time passes, the symptoms of dementia can be challenging and life-changing, requiring full-time care and attention, often provided at care homes. Apart from medical treatments, there are also some non-drug therapies which can help control the symptoms.

For PwD who stay in care homes, it is found that interactions with other residents are limited and needs to be initiated by the caregiver [14]. However, due to the escalating demands on caregivers within care facilities, achieving this may not be possible consistently [15]. Thus, we look into exploring whether a SAR can support caregivers by performing group activities with the residents, and keeping them occupied and engaged. This provides caregivers the time and space required to focus on personal needs of each resident, thereby reducing their workload and improving the quality of care.

Although the primary focus of this research is early-stage dementia individuals residing in care homes, we opt not to conduct experiments directly involving them. Instead, we perform trials with older adults aged 60 and above, as dementia primarily affects individuals within this age bracket [16]. This precaution is taken due to the vulnerability of the original audience. Our current objective is to gather feedback during this exploratory stage of the study. Post this, the feedback can be utilized to improve the system and subsequently, trials can be performed with individuals with dementia.

We study the effects of HRI activities done in groups to understand whether they can enhance interaction and thus reduce loneliness. Music therapy exercises that can stimulate brain activity and improve attention are applied. Along with this, the research delves into performing real-time engagement to evaluate whether the participants are engaged in the activity, and carry out re-engagement techniques to get them involved.

1.3. Motivation

Humans are social beings who rely on support from friends and family for survival and well-being. In literature, social engagement refers to the nature of interaction and behaviour of individuals within a group [17][18]. It defines the nature of bonds and connections within groups, describing how members of a group interact with each other [19]. Research shows that a higher level of social engagement is associated with better cognitive function and hence, an improved quality of life [20][21].

Studies show that the level of interaction among PwD residing in a care home is limited [22], leading to lower level of social engagement. Thus, living in a care home can increase loneliness, especially in elderly adults with dementia, and create a disconnect with peers [23], reducing the quality of life. Participating in group activities can alleviate the loneliness and promote social interactions [24], which can lead to better health and quality of life [25]. Though caregivers in nursing homes often organize group activities like singing, baking and games [24][26], their busy schedule can limit the frequency of such events [26][27][28].

Here, we propose the integration of robots in nursing homes to conduct these group activities. While the caregivers attend to routine tasks, a robot can actively engage the residents, potentially cultivating social interactions. Though limited, studies show that when a group of individuals interacts with a robot, human-human interactions increase [29]. This could lead to increased social connections for residents of care homes, and reduced loneliness.

This thesis specifically examines the use of a robot to lead sing-along sessions with groups of residents. This activity is chosen owing to the extensive benefits of music therapy as a non-drug therapy. This is further elaborated in Section 2.1. Since the purpose of the robot is creating an emotional connect for PwD with peers, and even possibly the robot itself, it is categorized as a SAR.

1.4. Research Gaps and Areas of Research

In this section, we identify certain gaps in the current research surrounding HRI with groups of people, and subsequently define the main topics explored within this thesis.

1.4.1. Real-time engagement detection

As mentioned in Section 1.1, the primary attribute of humanoid SARs is the possession of human-like characteristics, one of which is the ability to identify whether the person is enjoying a conversation with someone.

Identifying user engagement requires social understanding, and could be subjective and complex even for humans. It involves many subtleties like body language and eye gaze [30], which can be especially hard to interpret in a group environment. Thus, equipping a robot with engagement detection capabilities becomes challenging, as it is difficult to model such a complex characteristic with rules. While many solutions are proposed to incorporate this in robots, they all have limitations in the context of the thesis, which will be elaborated in Section 2.3. Thus, engagement detection, especially in a group of older adults, is a niche area of study [31], requiring extensive research and exploration.

In this thesis, we look into contextually tailored methods that a robot can use for detecting engagement of users, with the goal of gaining insight into users' perceptions of this situationally aware robot. Most engagement detection techniques in HRI do not have an active role for the robot. In other words, participation engagement is detected in the background using algorithms and the robot only participates in re-engaging the participants. We offer a fresh perspective to this by deviating from mainstream passive techniques of engagement detection and leverage the robot's presence to understand the involvement of the user in the interaction with the robot. This is further explained in detail in Section 3.2.

1.4.2. Group dynamics among care home residents

While designing for a specific group of people, we must be aware of their behaviour, experiences and preferences. Since our user group is persons with dementia living in care homes, it is imperative to understand the dynamics within groups, social interactions and communication patterns. This helps us design the behaviour of the robot to better meet their needs, ensuring that the intervention contributes positively to their well-being. For this, we research relevant papers and interviews that have been conducted with healthcare personnel that give an insight into this area. This is elaborated in Section 2.5.

Understanding the target audience allows us to effectively design a robot for group HRI. Currently, most of the research around group HRI focuses on how human-robot interaction and the role of the robot. However, there is a noticeable gap in research concerning *human-human* interaction in the presence of a robot [30]. Specifically, there is limited exploration examining how group interaction with a SAR influences communication among peers and the sense of belonging within the group. This thesis looks to provide insights into this specialized area. We particularly look to investigate whether participating in a robot assisted sing-along session *as a group* can improve human-human communication in a care home.

1.5. Research Questions

Due to the vulnerability of people with dementia in trials, the initial research focuses on older adults, specifically in the range of 60 to 80 years. This demographic can be considered an emulation of the original target group, since dementia mostly affects individuals in this age bracket [16].
Keeping this in mind, the research questions are presented below.

RQ1: How do older adults perceive a robot which moderates a sing-along activity, with real-time engagement detection and re-engagement strategies?

This research question delves into the acceptance of SARs in the community of older adults. We aim to explore how they respond to the robot and react to it when it detects their engagement level and attempts to motivate and re-engage them. With this, we try to understand the comfort level and concerns associated with human-robot interaction. This could provide valuable feedback which can inform the design choices and future iterations of such a robot, ultimately increasing its usability and effectiveness in promoting social engagement among older populations.

RQ2: Can participation in a sing-along session moderated by a robot with a group of older adults enhance interaction with peers?

While group activities definitely have a benefit in enhancing social engagement and experience, we seek to understand whether one moderated by a robot can achieve the same effects. The session is specially curated to incorporate opportunities for interaction among the participants. We explore the extent of social connection experienced by the group as they interact with the robot and with each other during the session. Communication and increased engagement with peers is beneficial, especially for people with dementia in a care home, as it provides a sense of belonging and trust among peers, thereby possibly improving quality of life.

1.6. Report outline

The report is structured into 6 chapters. Chapter 2 outlines the existing research and literature in various fields involved in this thesis. Chapter 3 describes the methodology used for implementation, including the robot used, the framework to program it, the flow of a session and the proposed technique for engagement detection. Chapter 4 details the experimental methodology, its design and evaluation details, including details of the participants and measures used for evaluation. Chapter 5 presents the findings of the conducted experiments, followed by a discussion about the results and the generalizations that can be made. These observations are used to answer the research questions, and finally make design recommendations for future work. Chapter 6 concludes and states the limitations of the current work and proposes directions for future work.

2

Literature Review

In view of the research question, this study utilizes and touches upon various topics, which can be categorized into the following themes: music therapy, HRI with older adults, engagement detection, adaption of robots, group dynamics in care homes, group HRI and a metric review in HRI. We summarize the findings of the studies conducted in the aforementioned areas of research in the following sections.

2.1. Music Therapy

A systematic review of randomized controlled trials conducted in the field of music therapy for individuals with Alzheimers [32], reveals multiple studies that indicate a positive effect on cognitive function. Active and passive techniques of music interventions are used in therapy [33]. Active interventions entail active partcipation from the user, such as singing along with the music, whereas in passive interventions, participants are more receptive, and only listen to the music. It is also found to fuel memories and enhance relaxation among users [33]. Familiar music can also arouse episodic memories and activate neurons in the brain, which otherwise could remain dormant in individuals with cognitive impairments [34].

A study carried out in the UK involved singing sessions with 10 patient-carer pairs [35]. Interviews with carers showed numerous benefits of singing in a group for PwD. They note that regardless of the stage of dementia, they are seen to actively participate and enjoy themselves. The activity improved the bonds between PwD and their caregivers, providing an opportunity for them to interact and engage. It also offered a platform for caregivers to meet each other, giving a sense of community and belongingness. Moreover, the activity stimulated memories in PwD, and helped in coping with and accepting dementia.

Multiple benefits of Karaoke training as well are found in older adults, including but not limited to improvement in cognitive skills, tongue pressure and respiratory function [36]. Following a 12-week intervention wherein the control group performed scratch art and the experimental group performed Karaoke training, a significant difference was found in the Frontal Assessment Battery at bedside (FAB) test to assess frontal lobe function [37].

Another study done with people with Alzheimer's disease showed preliminary results of the usefulness of singing training in improving neural efficacy of cognitive processing [38]. The intervention was 6 months long, after which the time to complete the Japanese Raven Colored Progressive Matrices (RCPM) was significantly reduced, showing an improvement in psychomotor speed. It also showed a significant improvement in the sleep time of those affected with Alzheimer's and a significant decrease in the Neuropsychiatric Inventory (NPI) score.

A qualitative evaluation of community singing for youths with mental health concerns showed numerous benefits of music therapy in groups [39]. A setting, called Coffee House, invited adolescents and staff members of the mental health care facility to join. Here, participants performed together and the primary goal was to involve maximum number of people. Although the target group in this analysis was not individuals with dementia, it highlighted some important advantages of music in a community. Results showed that participants felt accomplished and included. It also enhanced their self efficacy, diminished self consciousness, helped them overcome anxiety and most importantly, improved their social connect with peers.

The benefits of music therapy are evident and thus, we propose a robot guides users through a sing-along activity session. This serves as a non invasive therapeutic activity which can be advantageous for individuals with dementia.

2.2. HRI with elderly and cognitively impaired individuals

Robots have been used in various forms for the community of older adults. In one instance, the Misty Robot was designed to engage with older adults in unstructured conversation [40]. A Wizard of Oz approach was used to investigate whether varying the level of interactivity of the robot made a significant difference in users' experiences. Although the level of interactivity showed no significant differences in users' perception of the robot, a significant improvement in the development of friendship between the user and robot was noted over time.

The Danish Care System for elderly has also incorporated the use of robots for various tasks including cleaning, entertainment and assisted therapy [41] to reduce workload on staff and improve quality of life. It was found that the general attitude towards robot technology is positive with very little resistance. However, the expected standard of robots is quite high: they must be stable and easy to use to be integrated into a care home.

A 10 week intervention was conducted in a care home for older adults with the Pepper robot to understand barriers and enablers of robot systems [42]. The users were first given a chance to get accustomed to the robot with a music related game, after which it provided physical and cognitive exercises. It was observed that it was essential for participants to familiarize themselves with the system to become less reserved and participate actively. Accessibility issues like loss of hearing did create some difficulties. However, an overall positive attitude towards Pepper was noted. Sometimes, the robot was found to be awkward, whereas other times it was friendly.

A robotic architecture system was designed to use a SAR to engage pairs of older adults, both with and without cognitive impairment, residing in assisted living facilities [43]. Different activities and games which stimulated human-human interaction were conducted, while participant engagement was monitored. It was found that the nature of human interaction is influenced by the type of activity and the specific form of engagement it entails, whether verbal or visual, among participants. For example, activities that required users to make eye contact or engage in conversation with each other led to heightened levels of social interaction.

An experiment used the Ka Ka robot, which explored the effect of humanoid social robots at home, with older adults [44]. The robot had four features, namely interactive modalities, calendar planning and task reminders, promoting a healthy lifestyle and puzzle games. This aim was to offer companionship to the older adults, and stimulate their cognitive functioning. A qualitative analysis found that the robot provided emotional support, diversified their activities and enhanced the dyadic relationship between the older adults and their families.

Another trial conducted with individuals with dementia used the Nao humanoid robot [45]. Four types of sessions were designed by specialized therapists, which aimed to cognitively stimulate the users: storytelling, music therapy, language and physiotherapy. These activities were focused on persons with moderate dementia, and showed a trend of improvement in neuropsychiatric symptoms, although no statistically significant difference was observed. However, preliminary results showed promise as the robot captivated the attention of the elderly, thus laying the groundwork for future research in the field.

Additionally, a review was conducted in the field of assistive social robots in elderly care [46], which summarized existing literature from various data sources in the field. This review paper highlights key findings of studies involving pet-like companion robots, whose primary role is the psychological well-being of the user. Most studies report an increase in the positive mood of the user as evaluated from facial expressions. The elderly people are also reported to be less lonely and more stress free. Social ties amongst the elderly in care homes are found to be enhanced, as well as the relation between the elderly and their families. However, it is important to note that majority of these trials were carried out in Japan and limited to the use of specific robots like Aibo, Paro, iCat and Pearl. Nevertheless, these studies lay a crucial foundation for future large scale and cross cultural research in the area.

Thus, despite being a relatively new and upcoming field, there is sufficient literature to show the benefits of HRI in the elderly population. Specifically, HRI holds immense potential for enhancing the well-being and mental health of these individuals. This motivates us to advance in this field, with designs tailored for cognitively impaired adults, aimed at providing companionship, increasing socialization and mitigating loneliness.

2.3. Engagement Detection in HRI

Engagement detection is an inherently human trait, which gives us the ability to assess a person's level of involvement during a social interaction. Humans typically recognize subtle clues from body language, tone and facial expressions, and accordingly react and adapt during the interaction. It is a social skill which human beings typically develop instinctively. Thus, while considering the design of human-like SARs, incorporating this skill becomes increasingly important.

There are various different approaches taken in literature to integrate engagement detection in HRI, which can be organized into these different, although not mutually exclusive, categories, discussed below.

2.3.1. Emotion Recognition

The most intuitive approach to detect whether a user is enjoying an activity is by observing their facial features and expressions to understand their emotional state. A framework proposed for real time engagement detection utilizes Computer Vision (CV) techniques to extract non verbal features and use them to detect engagement using Random Forests [47]. Intel's pre trained models (OpenVINO toolkit[1]) are used to extract the user's facial expressions, head pose, gaze direction and distance variation between the camera and user. These features are used for training the engagement detection model and using it to eventually make a prediction of the user's engagement level. Training is done using the DAiSEE dataset [48] which consists of several user videos and four corresponding outputs: boredom, engagement, confusion and frustration characterized into 4 levels of intensity. Thus, features are extracted from these videos of the dataset, and a Random Forest based model is trained on the features and engagement level outputs. Subsequently, this model is used to predict and classify the level of user engagement in real time.

A pilot study was also done to compare an empathic version of a SAR with a non empathic version [49]. The empathic version of the robot utilized facial expression and speech sentiment to detect the user's emotional state, and accordingly utilized an affective dialogue manager to generate a response. The empathic version showed significant improvement in the reported mood of the user, as compared to the non-empathic version. The emotional state is modelled as a multimodal variable, using an equally weighted sum of facial and verbal sentiments. The Facial Expression Recognition (FER) is performed using a Deep Neural Network (DNN) model [50], which returns the probabilities of positive, negative and neutral emotions. For verbal sentiment analysis, CoreNLP [51] was used, which returned the emotion of the user based on the speech content.

An evaluation was performed in education, with the Nao robot which provided tailored support during a quiz when the student's negative emotion was detected [52]. During the interaction, the robot took pictures of the users, which were then sent to a server, where emotion recognition was performed. The DeepFace attribute analyzer [53] was used on the server, which could recognize the six basic emotions: anger, fear, happiness, sadness, surprise and disgust. Recognizing the limitations of this approach, the authors acknowledge that the photos are highly dependent on factors such as lighting conditions and camera angles. Thus, while this may work in a controlled lab setting, making it work in a real life situation poses a challenge.

Though limited [31][54], there has also been research on emotion detection in groups. One such study [55] uses the Viola Jones algorithm [56] to detect all the faces in a frame, and then the VGG-16 neural network (VGGFace) [57] to detect one of the six basic emotions in them. These emotions are then aggregated over multiple scenes and frames to obtain the final emotion of the group. Notably, there are drawbacks to this approach similar to [52] such as low accuracy in a real life scenario and presence of too many people, as well as high reliance on lighting.

Though widely used, the approach of emotion recognition should be taken with a pinch of salt and should be approached with caution as pre-trained models might not always work [58]. A survey highlights four critical lessons while perceiving and sensing human affective behaviour during interactions [58]. Firstly, the six basic emotions are not always relevant, and depend on the context. Furthermore, it usually only works in a controlled environment with no variations in light, as also pointed out in [49][55]. Secondly, facial expressions might not always be a relevant signal for affect recognition, and affect recognition accuracy is not the sole goal. Thirdly, pre-trained models are not generalized across contexts. For example, since most of the training data consists of middle age adults, it might not work for other age groups like older adults. Finally, in a situation which requires adaptation and personalization, emotion recognition alone could be insufficient.

[1] OpenVINO toolkit: Intel's pretrained models

2.3.2. Pose and Gaze Analysis

A way to determine engagement is also to observe the direction in which the user faces and their body language. To this effect, a solution was proposed which used Recurrent Neural Networks (RNN) to detect user disengagement during interaction with the Pepper robot [59]. The researchers made use of the UE-HRI dataset [60] which consisted of videos of spontaneous interactions between humans and the Pepper robot. Segments of these videos where the user engagement decreased were manually annotated based on verbal and non-verbal cues. Social signals of distance from the robot, gaze direction, head pose and speech were extracted using the robot sensors coupled with the NAOqi-SDK [2]. A deep learning model was then trained on these features to predict user disengagement in real time.

Another study done aims to measure the "with-me-ness" or the attention focus of humans using head pose [61]. A fast template-based face alignment algorithm [62] implemented in the *dlib* [63] library is used to extract 68 facial features. 8 of these features are then matched to the 3D counterparts. OpenCV's iteratve PnP algorithm [3] is used to find the angle or rotation of the head with respect to the camera. The field of attention, or where the user is focusing their gaze, is then considered as an aperture of 40° from the human view. In this way, the objects that currently have the user's attention can be assessed. This method can be used to evaluate whether the user is looking at the robot and is involved in the interaction.

An evaluation done also analyzes static features like focus of attention, gaze and head pose, along with dynamic features which are measured over time, such as response time and imitation [64]. The constrained local model approach [65] is used to track faces and find the head pose. This method along with data mining is used to assess other static and dynamic features and accordingly measure engagement of the user.

In a group environment, where multiple people are interacting simultaneously, it becomes tricky to evaluate engagement and enjoyment based on gaze direction alone. Additionally, in such a scenario, factors such as shared attention and overlapping gazes further complicate this analysis.

2.3.3. Deep Neural Networks Trained on datasets

A number of researches customize their model based on their specific use case and context. This provides for higher accuracy and precision, since they are tailored to domain specific features. The general outline for such an approach includes collecting data, manually annotating it by assigning labels, training a Deep Neural Network on the videos and their corresponding labels, and finally employing this model to predict engagement in real time.

One such research first collects data from an autonomous tour guide robot (TOGURO) deployed in a museum [66]. The interaction between the user and the robot is recorded, and later manually annotated to establish ground truth. These manual annotations, or labels, can have three values signifying the level of engagement of the user during the interaction: HIGH, MEDIUM and LOW. The network architecture used to train the model comprises of two parts: a convolutional module to extract features from each frame, and a recurrent module to aggregate the frames and produce a temporal feature vector. This model is further tested on the UE-HRI dataset [60] to assess generalization ability, and gives encouraging results.

A field study done explores the functions of an empathic robot in a group learning scenario [67] in the context of a game. To design a robot with empathic competencies, multiple steps are involved. Data is first collected using mock up studies with potential end users, i.e, students. These students interact with human teachers, which is recorded. Then, a Restricted Perception Wizard-of-Oz (WoZ) is employed where experts, who have access only to processed observations from the interaction, remotely control the robots. These inputs include perceptual features such as facial features, auditory features and game related data. Data from the Restricted Perception (WoZ) is used to create the Hybrid Behavior Controller, which outputs animations, gaze and dialogue based on the perceptual features of the participant.

Although the employment of use case specific datasets is beneficial due to the utilization of context specific clues, the collection and annotation of these datasets can be an arduous task, making it infeasible in many situations. Moreover, datasets for specific use cases are not readily available. Additionally, in the context of this thesis, accurately capturing perceptual features in a group setting and training a model could present difficulties.

[2] NAOqi API
[3] OpenCV Iterative PnP

2.3.4. Wizard of Oz

Wizard of Oz (WoZ) refers to an experimental technique used in HRI, which involves a human operating a robot behind the scenes [68]. The participants are under the impression that they are interacting with a robot, but in reality, it is being operated by a human controller. This method is used in the initial stages of a project, and allows designers to gather feedback about required functionalities, before automating it.

A study done looks to evaluate the use of robots as a social companion for elderly people [69]. Three experiments are performed, one of which uses the Nao robot [4] with a WoZ approach. The robot engages the user with humour, and expresses empathy to build a deeper relationship. [70] presents an open source WoZ interface for the Pepper robot, which can allow researchers for a quick hypothesis testing. The interface has several configurable features, allowing a human to easily control the robot's actions and speech. This makes it easier for non technical researchers to conduct experiments with the Pepper robot.

A systematic review conducted for WoZ studies in HRI showed six major areas where this approach is employed: Natural Language Processing, Non verbal behaviour, Navigation and Mobility, Manipulation, Sensing, Mapping and/or localization [71]. This also highlights the suggested guidelines that should be followed while conducting experiments through WoZ [72]. The three requirements that must be met while considering a WoZ approach are: simulation of the future system must be possible given human limitations, it must be possible to specify the future system's behaviour and it must be possible to make the simulation convincing.

Therefore, a WoZ approach could be very useful in an initial analysis, in the requirements gathering phase, and to obtain feedback. Given the exploratory nature of this study, we integrate the WoZ technique with an automated system to monitor and respond to low levels of engagement. Further details are explained in Section 3.2.

2.4. Re-Engagement Strategies in HRI

When disengagement is detected by robots, different measures can be implemented to re-engage participants. A user study done to evaluate the kind of robot that elicits social responses [17] found that robots with human-like characteristics and even minimal social cues were more engaging than those with none. Gestures, facial expressions and lip movements in addition to the voice of the robot were found to invoke a higher level of engagement in the participants.

A review of the studies of engagement in HRI [31] found that the use of gestures, direction of gaze, nodding and volume modulation can help gain attention of members. [73] also suggests that gestures and engaging in conversation regularly are beneficial strategies for sustaining attention. Additionally, addressing participants by name also plays a significant role in enhancing engagement levels.

An intervention utilizing SARs to promote social engagement among older adults [43] shares a comparable use case with the current thesis. Activities emphasizing social, physical, and cognitive skills were conducted to benefit the users. The study coupled verbal encouragement with addressing participants by their names to motivate their participation. Throughout these activities, the robot utilized various gestures such as clapping, raising its arms, and waving to promote both human-robot interaction and human-human interaction.

2.5. Group Dynamics in Care Homes

A mixed method systematic review [14] compiles and evaluates various studies conducted to survey special care units for dementia. This review paper primarily addresses the relationship among the residents of care homes. Studies show that social interaction can be a measure of quality of life (QOL); however, the level of interaction among residents of a care home is sadly very limited [22]. Majority of the time, residents are not seen actively conversing. However, opportunities for interaction do exist: mainly in communal spaces like dining rooms and activity areas [22][74]. Studies suggest that most interaction takes place in groups of four to six, and the most conducive time for the same are mealtimes [74][75].

Group activities in care homes are found to provide a feeling of community and have a therapeutic impact on residents [76]. A study found improvement in mood and participation of people with dementia who engaged in group activities [24]. 10 different activities including exercise, choir singing, baking, games, etc. were chosen based on therapeutic recreation staff expertise. Using the Group Observational Measurement of Engagement (GOME)[77], significant improvement in mood was found as compared to the control group which did not perform group activities.

[4] Nao Robot

Research is also done to explore the dynamics of interaction and relationships between residents and caregivers in nursing homes. Surveys note that fostering connections in this setting can be a challenge due to physical and cognitive weaknesses among members [15]. It is observed that one of the factors that affects the relationships with caregivers is the consistency of staff. Maintaining the same caregiving team creates a sense of continuity in the lives of residents, and allows for a deeper connection [15][26].

Interviews with 15 healthcare personnel from special-care dementia units provide information on the social community of residents [26]. As found in other studies [15], familiarity of caregivers can go a long way in forming a strong connection with residents. Constant rotation and changes in staff can adversely affect resident-staff bonding. Individuals with dementia also need constant reassurance and need to be reminded that they are in a safe space. Additionally, physical presence of personnel is important, which might not be possible on busy days.

This suggests that integrating robots into care homes can be beneficial. They offer consistency, have a physical presence, and can establish familiarity over a long period. On busy days for staff, they can engage with residents. Robots can also be used to conduct group activities to enhance social interaction - a focal area of research within this thesis. However, it is important to note that robots cannot replace healthcare personnel in any capacity, but are merely a supplement or an adjunct to the care home environment.

2.6. HRI in Groups

Group HRI involves the study of interactions involving multiple individuals with one or more robots. It investigates how users and robots communicate and collaborate with each other, along with the dynamics and challenges that arise during this interaction. It is different from the traditional one-on-one HRI interaction and is crucial for developing robots that can integrate into society.

A study makes use of this concept within pairs of children, where a robot serves as a mediator during a play session [78]. If conflicts arise, the robot steps in and offers prompts for constructive conflict resolution. This helps the development of interpersonal conflict resolution skills in children. Interactions were controlled using the the WoZ technique. Along similar lines, another evaluation studied the use of a robot for conflict resolution in a group of adults during a problem solving task [79]. Research also delves into the use of a robot as a moderator within group interactions during a tablet-based assembly game [80].

A literature review done in the field of group HRI summarizes and highlights several findings of interactions between robots and teams of people [29]. It is found that groups, as opposed to individuals, are more likely to interact and engage with a robot. Moreover, when groups interact with a robot, the human-human interactions increase. Individuals are found to spend more time conversing with each other rather than interacting directly with the robot. Additionally, the expression of emotional states seems to change significantly as compared to one-on-one interactions.

Thus, interaction in a group is particularly complex as not everyone may participate with the same interest. One can address either the whole group or just one person, which can can make it difficult for the robot to understand and capture the dynamics [31][54]. Moreover, there are very few metrics to assess functioning of groups in HRI. Apart from this, we overlook factors like robots being deployed in a lab setting instead of a real life context, and fail to account for cultural differences between participants [54]. These factors are not captured in traditional questionnaires, making it difficult to generalize results. Thus, a combination of different types of methods must be used to study one particular subject.

These reasons might potentially explain why this field is still in its nascent stages, with much left to explore. Though there are a few studies which focus on the dynamics of human-human interaction with a SAR [43], most research focuses on the role of the robot, and how users can better converse with it [30]. This thesis, thus, aims to contribute to the niche field of group HRI focusing on interactions among older adults. Keeping in mind the challenges of this domain, we opt for a qualitative and quantitative analysis to gain a deeper understanding of the findings.

2.7. Evaluation Metrics in HRI

Human Robot Interactions can be evaluated in multiple ways, using surveys, questionnaires, interviews and qualitative or subjective studies [81]. While questionnaires are used to obtain answers to specific questions and hypothesis driven studies, qualitative studies are utilized for exploratory purposes and data collection.

A review done summarizes the psychometrically validated, frequently used questionnaires in the field of HRI [82] to evaluate user perceptions towards robots. The most highly cited questionnaire is the Negative Attitude towards Robots Scale (NARS) [83] which consists of 14 items that need to be rated on a five-point Likert scale. As the name suggests, it addresses anxieties and apprehensions that people might have about robots. Thus, for most questions, a higher item score indicated a more negative attitude. The second questionnaire, Robotic Social Attributes Scale (RoSAS) [84] was developed from the Godspeed questionnaire [85]. It consists of 18 robot characteristics which users must rate on a nine-point Likert scale, depending on whether they think the characteristic is associated with the robot. Aiming at three broad subscales - warmth, competence and discomfort, it gives an overall idea of the users' perception of the robot. Another scale, the Ethical Acceptability Scale [86] is used typically used to assess the ethical ramifications of an experiment, specifically the use of robot-enhanced therapy for children with autism. It consists of twelve items to be rated on a five-point Likert scale. The Technology Specific Expectationss Scale (TSES) [87] is used to assess expectations of users prior to the interaction with a robot. This is because the satisfaction of a user depends on the extent of their initial expectation, followed by their experience with it. It consists of 10 questions on a five-point Likert scale. The Frankenstein Syndrome Questionnaire [88] also measures the acceptance of humanoid robots including expectations and anxieties towards it. It contains 30 items on a seven-point Likert scale, and was developed by exploring cross cultural factors. Finally, the Multi-Dimensional Robot Attitude Scale [89] attempts to create a measure that can capture a larger range of aspects related to the attitude of the robot. It was developed using a cross national sample across China, and consists of 49 items with a seven-point Likert scale.

On the other hand, qualitative studies can be a very valuable tool to gain insights during the initial phases, and are usually used in an exploratory study. A survey of qualitative methods [90] shows that most qualitative studies in HRI can be divided into six methods: qualitative observation, semi-structured interviews, focus groups, generative activities, reflective and narrative accounts, and textual/content analysis. Qualitative observation studies help researchers understand the non-verbal behavioural aspects, whereas interviews enable participants to articulate their experiences in their own words. Focus groups facilitate the collection of large amounts of data in a short time, and aid in diverse social knowledge and varying perspectives in a group setting. While reflective and narrative accounts use creative descriptions from participants to obtain subtle insights into the design, textual analysis allows users to provide free form answers to structured questionnaires.

Selecting the appropriate metric for evaluation depends on various factors, including the context, the objective of the study, the phase of the research and the participants involved. Since this study is in the initial phases of exploration in the field, we focus on gaining user insights about their experience, and understanding their attitude towards the robot. To this effect, we employ a blend of qualitative and quantitative studies for a well rounded and balanced assessment. This is explained further in Section 4.4.

2.8. Discussion

We have seen numerous benefits of music therapy and singing, especially in PwD, along with the vital role played by group activities in building bonds with peers in a care home. HRI with older adults and PwD has also shown to have extensive potential in improving their mental health. Interviews with healthcare personnel provide insights into the life of residents in care homes, and we note that while the staff attempts to organize activities for them, it might not always be possible. Thus, we aim to design a system which enables robots to interact with groups of people and guide them through an activity, specifically chosen in the realm of music therapy. This not only aims to enhance social bonds among the residents of the care home, but also seeks to reduce the workload of the caregivers.

We also looked at various techniques of engagement detection that can be incorporated in the robot to discern disengaged users and motivate them. The most common ones are emotion recognition, head pose estimation, training deep neural networks and WoZ. However, all these methods have limitations in the current context. Emotion recognition techniques rely on light sources and distance from the camera, while head pose estimation methods will not always work in a group setting. Collecting data for the precise use case and training deep neural networks on the data takes time and raises privacy concerns. Hence, it is out of the scope of this thesis. While WoZ is a good start for an exploratory study, we must keep in mind that it should be suitable for automation in the future. Keeping these limitations in mind, we propose a new method of engagement detection, curated for a group sing-along activity.

After exploring the existing research in the domain of group HRI and PwD, we create a robot prototype to address the unique requirements of the target audience in the context of this thesis. With this prototype, experiments are conducted and feedback is obtained, which can aid us in answering the research questions.

3

Prototype Design and Implementation Methodology

The objective of the thesis is to design a robot which can guide a group of individuals through a sing-along activity, while performing engagement detection in real-time. In this chapter, we lay out the structure of a typical interaction session with the robot, delving into the step-by-step flow of the activity. A detailed explanation of the proposed engagement detection technique employed during the session is provided. After the desired prototype has been established, we discuss the hardware and software components used in developing it. The technical details of the robot and the rationale behind the choice of the robot is explained. Various frameworks used for implementation are presented, along with a discussion on the frameworks chosen for this implementation. Following the discussion of the design in this chapter, the prototype undergoes testing through experiments involving older adults, the details of which are presented in Chapter 5.

3.1. Robot Implementation Flow

This section provides a comprehensive overview of what one session with the robot constitutes. It starts with the set-up required before the start of the experiment, including the robot and peripherals. The stages of the expected interaction between the robot and the users are outlined, elaborating on the conversation flow and other key elements involved. With each stage, we give examples of the expected dialogue sequence within that stage. The names 'Jane' and 'Jack' are used as placeholders for participant names in the examples given. These are replaced with actual participant names during the session. The entire list of dialogues is included in Appendix A.

It is important to note that all interactions in the experiments conducted with the robot are in Dutch. This language choice is based on the fact that all participants are older adults recruited from The Netherlands, who might not always be fluent and comfortable speaking in English. Interactions with the robot in their native language seek to make them feel more comfortable and ensure clear and effective communication. However, for coherence in the report, we provide examples of interaction in English.

The flow of a typical session is shown in Figure 3.1.

Step 0: Set-up

Before the start of the experiment with participants, the set-up involves multiple tasks with the connection of different devices. This can be visualized in Figure 3.2.

- Pepper Robot: The Pepper robot is used in the experiment, the details and design rationale for which is explained in Section 3.3. At the core of the experiment, the robot is switched on and connected to the two frameworks utilized to program it, SIC and Robots in de Klas, as discussed in Section 3.4.

- Philips Television Screen[1]: This serves to display the different options that users are presented with to choose their desired song, as well as the lyrics of the currently playing song. An alternative to using this screen was to use the tablet attached on the chest of the Pepper robot. However, due to its small size

[1] Philips Smart TV

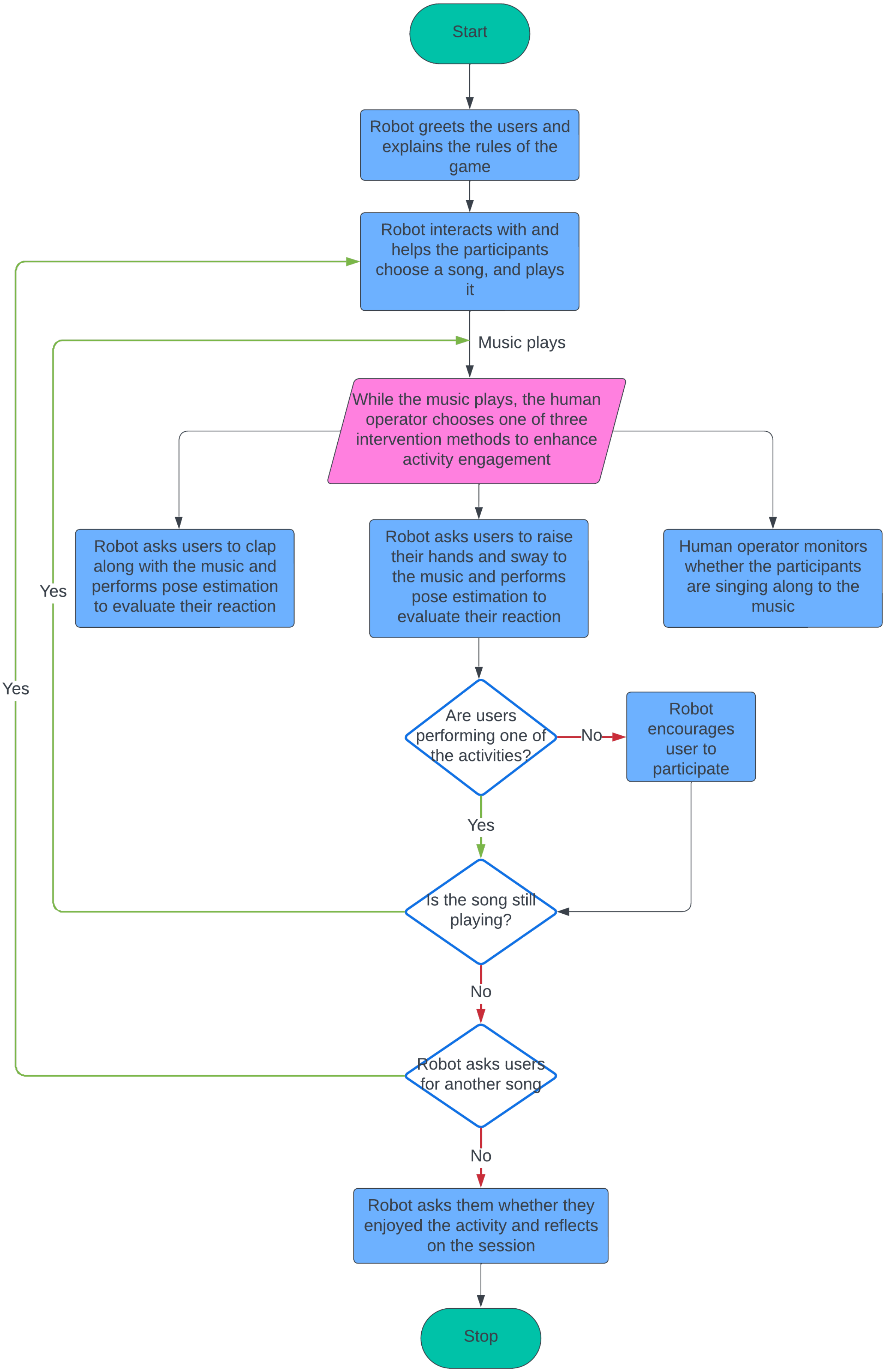


Figure 3.1: Proposed Flow of the Experiment

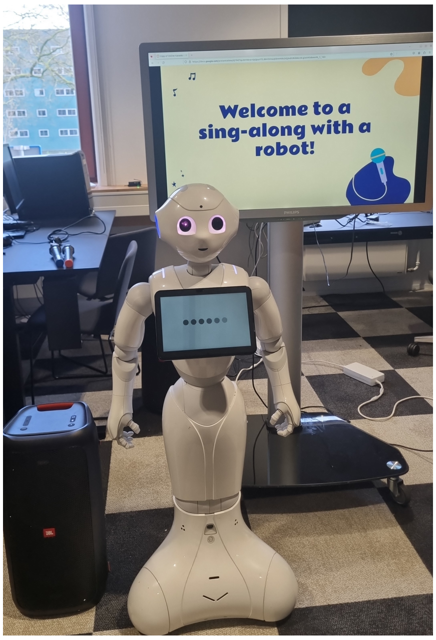


Figure 3.2: Experimental Setup

and potential visibility issues for participants, we have chosen a 42-inch screen to accommodate larger font sizes. The screen is connected to the laptop via an HDMI cable, and songs are played through Spotify[2].

- JBL PartyBox 100[3]: Two speakers are used as the central sources for playing the chosen song, contributing to creating an immersive and energizing environment for participants. The use of these speakers is pivotal in recreating the ambiance of an authentic music room, thereby enhancing excitement in the users. They are connected via Bluetooth to the laptop.
- Lenco Wireless Microphone[4]: Four wireless microphones are connected to both the JBL PartyBox 100 speakers via a receiver, to amplify the users' voices while they sing. The microphones, along with the speaker contribute to creating a lively and exciting atmosphere for the users.
- Logitech Webcam[5]: Connected via a USB cable to the laptop, the webcam serves to capture participant video in real-time, to monitor pose and thereby estimate engagement. Initially, the use of the camera mounted on the Pepper robot was attempted for this purpose. However, the camera did not provide the required clarity. Additionally, the camera's stability was compromised by the movement of Pepper's head, leading to inconsistency in the captured footage. Thus, we use an external camera for this purpose.
- Samsung Galaxy S21 FE Phone[6]: Participants and their interactions with the robot are recorded with the Samsung Galaxy S21 FE. The phone was calibrated to include all the participants in the video recording. The robot and the television screen are not captured.
- Names of Participants: Participant names are collected and input into the code to make the interaction more personalized for users. The names are deleted from the code after the experiment, in compliance with General Data Protection Regulation (GDPR).

Step 1: Introduction

The experiment involves 3-4 participants seated in front of the Pepper robot. The robot starts by greeting each participant individually by their names, thereby creating a personalized and immersive experience. Pepper

[2] Spotify
[3] JBL PartyBox 100
[4] Lenco Universal Dual Wireless Microphone with Receiver
[5] Logitech C920 Pro HD Webcam
[6] Samsung S21 FE

then introduces itself and outlines what the session is going to be about. It explains that it will play their favorite music, and the participants are encouraged to sing along and enjoy themselves. During the greeting, waving gestures are employed to make it more human-like, also engaging the users [17].
Examples of Dialogues:
Robot: *(waves arms to say hi)* “Hi Jack and Jane! My name is Pepper. I hope you are having a good day. Would you like to play a game?”
Robot: “I will explain the rules of the game. I have many songs to choose from, and I will help you make the choice too. Once you all have chosen a song, let me know and I’ll play it for you. You then sing along to your favorite music and have fun!”

Step 2: Music Choices

The robot actively participates in guiding the group toward selecting a song suitable for all participants. Instead of just giving song choices, it initiates the process by inquiring about the preferred language of the song they would like to sing, offering two options: Dutch and English. These choices are prominently displayed on the television screen to assist users in making their selection. Participants then engage in discussion and convey their language preference to the robot. Subsequently, the human operator, acting as the Wizard of Oz, selects the chosen language on the screen. The robot proceeds by further inquiring about the genre of the preferred song, with the available options also displayed on the screen. The available genres are Pop, Rock, R&B Soul and Folk. Participants engage in discussion to determine their preferred genre and communicate their choice to the robot. The Wizard of Oz once again clicks on the selected option on the screen, triggering the opening of a specially curated Spotify playlist tailored to the users’ choices. Subsequently, the robot prompts the users to choose a song from the list displayed on the screen. Upon selection, the Wizard of Oz clicks on the desired song, initiating playback. The objective is to improve interaction between the users themselves and enhance communication with the robot as it guides them through making a decision as a group.
Examples of Dialogues:
Robot: “Do you want to sing a Dutch or English song? These choices are also displayed on the screen.”
Participant 1 (Jack): *(looks at other participants)* “What do you prefer?”
Participant 2 (Jane): *(looks at other participants)* “Let’s go ahead with Dutch!”

Step 3: Prompts and Re-engagement of participants

The music is now playing, and throughout the song, we attempt to determine the engagement level of participants with respect to both, the singing and the robot. They are considered disengaged if they do not respond to robot prompts of clapping along or raising their hands (Robot engagement) and are not actively singing along (Activity engagement). The technical details of how this is achieved are explained in detail in Section 3.2. If participants are identified as disengaged, the robot tries to re-engage them by addressing them by name and encouraging them to join in. Apart from addressing individuals separately, the robot may also address the group as a whole, in cases where they are singing together or participating as a group.

In addition to engagement detection, the robot executes subtle movements throughout the duration of the song, such as swaying hands, raising hands, and applauding. These movements are coordinated in real-time by the human operator, ensuring they are performed at appropriate moments during the song. For example, the robot could raise its hands in the air during the beat drop, or even when instructing users to do the same.
Examples of Dialogues:
Robot: *(raises arms in the air)* “Raise your hands and groove to the music”
Participant 1 (Jack): *(raises arms in the air and sways while singing)*
Participant 2 (Jane): *(does not raise arms)*
Robot: “Jane, feel free to join in and be part of the group!”
Robot: “I am glad you are enjoying, Jack!”

Participants: *(all participants singing along)*
Robot: “Wow! You all are amazing singers!”

Participant 1 (Jack): *(not singing)*
Robot: “Hey Jack! Come on and join in, we would love to hear your voice.”

Step 4: Next song

Once the current song is over, the robot asks the users if they would like to continue the activity and sing another song. Participants discuss and decide among themselves, and according to their response, the human operator provides input to the code, which makes the robot act accordingly. If the participants decide to sing another song, the flow goes back to Step 2. If not, the flow proceeds to the next step.
Examples of Dialogues:
Robot: "Thanks for singing along to this song! If you're having fun, would you like to play another song?"
Participant 1 (Jack): *(looks at the other participant)* "I am having fun, should we sing one more song?"
Participant 2 (Jane): "Let's go ahead!"
Participant 1 (Jack): "Hey Pepper! We would like to sing one more song."
Robot: "I'm happy to know you are enjoying the activity!"

Step 5: Closing of the session

Once all the participants have completed their songs, the robot thanks the users for their participation. It then proceeds to ask them how they felt about the activity and whether they enjoyed it. This enhances the human-like characteristics of the robot, and allows the user to share their feedback in an organic way. To make the conversation more fluid, the robot also reflects on the enthusiasm of the group. For example, if the group was lively and energetic, the robot compliments their energy and expresses its enjoyment. If the group is on the shy side, the robot acknowledges that it is challenging to open up, but offers encouragement that future interactions may be more comfortable.
Examples of Dialogues:
Robot: *(assuming the group was actively participating)* "Thank you, everyone, for participating. I enjoyed hearing you sing, you are certainly a fantastic and enthusiastic bunch of people. Did you have fun? How did you feel about singing along with me?"
Participant 1 (Jack): *(looks at the other participant)* "I definitely had fun! But, I wish you looked at me while talking to me."
Robot: *(assuming the group was actively participating)* "Thank you for your feedback, and for joining me today. I hope to see you again soon!"

This concludes the different steps involved in an experimental session with the participants and the robot. In the next section, we further elaborate on the proposed technique for engagement detection - one of the pivotal aspects of the robot design.

3.2. Engagement Detection and Re-Engagement Strategies

Engagement detection is an intrinsically human characteristic, guided by subtle clues, which could differ based on culture and background. While this poses challenges for integration into a robot, we hypothesize that a robot capable of identifying user involvement and accordingly providing encouragement and support, is more anthropomorphic and embraced by users. As discussed in Section 2.3, numerous approaches have been attempted, yet a definite methodology remains elusive.

In the field of engagement detection in Human-Robot Interaction (HRI), conventional techniques typically adopt a passive approach. This means that in these methods, the robot employs algorithms in the background to quantify the users' level of engagement. However, the robot itself does not actively contribute to assessing engagement. Here, we propose a novel technique that leverages the presence of a robot and provides it with an active role, enabling it to gauge user engagement. This approach involves the robot explicitly probing and prompting participants and gauging their responses to assess engagement. This proactive approach represents a shift from traditional passive methods, potentially leading to more effective and intuitive human-robot interactions.

In this thesis, we put forward an approach that integrates this novel active methodology with a passive one, providing a comprehensive technique that involves both aspects. Here, we first establish our definition of engagement in the context of this thesis, and then delve into to the technical details of its execution.

3.2.1. What is engagement?

The definition of engagement varies and is dependent on the context chosen [31][64]. In the field of education, engagement could refer to the level of attention and active participation of students [52][91]. In a social setting, it could be characterized by the quality and frequency of interactions with peers [43].

In the current context of HRI, we define engagement as the degree of involvement exhibited by the user, both in the ongoing activity (the sing-along) and the robot itself. We divide the definition into two separate components, because each captures a distinct aspect of user engagement, briefly explained below:

- **Activity Engagement**: In the present context, activity engagement is detected using a straightforward indicator - individuals are deemed engaged if they participate in singing along with the music. Participation in the activity itself can be a primary indicator of their interest and their mood. It can serve as a measure of the kind of behaviour the robot must exhibit towards the participant - should they be motivated to join in, or should the robot simply acknowledge and appreciate their excitement and enthusiasm? This represents a passive approach, wherein the robot evaluates engagement in the background and responds accordingly.

  Analysis of Activity Engagement: Apart from being a measure of involvement and enjoyment during the session, we also perform a behavioural analysis on the activity engagement *after* the session. This is done to identify and peruse additional engagement indicators such as social cues and nuances which provide insight into the level of participation in the activity. Non-verbal behaviours like eye gaze, body language and volume of singing are examined to get an understanding of intricate details such as the participant's emotional state, responsiveness to the activity and whether such an activity can enhance social interactions. This survey is done from the video data gathered during the experiments, results of which are explained in further chapters.

- **Robot Engagement**: This represents the active approach, where the robot provides prompts to users and monitors their response to these prompts. Participants are presented with one of two prompts: either to raise their hands or to clap along to the music. If the users perform these actions, they are considered to be engaged as they actively listen to the robot's instructions and participate in the interaction.

  Analysis of Robot Engagement: Responses to the robot can help us understand the acceptance of the robot into the community of older adults. Video recordings of the experiments conducted are scrutinized, similar to activity engagement, to gather more clues about the dynamics of the human-robot interaction. We look at reactions of users to the robot prompts, characterized by their facial expressions, eye gaze and their verbal comments. We investigate and address questions such as: Do participants respond to the robot prompts? If not, what are the underlying reasons? What are the factors that can be changed to elicit responses towards the robot? Finally, we judge whether the proposed method of detecting engagement is feasible, effective and efficient. These methodologies and corresponding results obtained are described in more detail in later chapters.

If the user is recognized to be disengaged, either in the activity, or the robot, it attempts to encourage the user to join in with the group. With the use of personalized comments, the robot aims to increase engagement while creating an inclusive atmosphere where everyone feels safe. After the session, we assess whether these comments made a difference and motivated participants to engage with the robot or in the activity. From the feedback and analysis of video data, we gauge the effectiveness of the interaction and motivation given and address questions like: Did the re-engagement strategies by the robot help? Did participants think the robot played a role in making them feel included? What factors contributed to the participants being re-engaged? With this, we make observations and recommend future designs that could improve it.

The implementation details of engagement detection and the re-engagement strategies employed by the robot are explained below. Activity engagement and subsequent re-engagement is performed using the technique described in Subsection 3.2.2, while the methodology for robot engagement is outlined in Subsection 3.2.3.

3.2.2. Activity Engagement: WoZ

Detection: The most intuitive way of detecting whether an individual is participating in the activity, is identifying whether they are singing along to the music - a visible indication of participation. This is implemented using a WoZ approach, wherein a human operator is present to manually observe and assess whether the participants are singing along to the music. This human operator serves as the "wizard" behind the scenes, actively monitoring participant behavior on behalf of the robot.

An automated approach was attempted, by tracking the amplitude of the input from each microphone, with each participant equipped with their own microphone. The objective was to gauge whether individuals were singing based on the audio signals from the mic. However, the mic picked up significant background noise including the music playing in the background and the robot's voice. Thus, isolating the participant's voice alone proved to be difficult and this approach was unreliable. Moreover, some shyer participants might prefer to sing without the microphone, rendering the automated approach less effective for capturing their voices accurately. As a result, a WoZ approach, involving human observation and judgment, was adopted instead.

Re-engagement Strategy: When the human observer notices that a specific participant is not singing, they input an encouraging message addressing the participant by name to the robot. If the entire group is singing enthusiastically, positive encouragement is given by addressing the group as a whole and complimenting their energy. This seamless process creates the illusion that the robot autonomously detects and responds to participant engagement, despite human intervention, thereby giving the impression that the robot is managing the entire interaction independently.

It is essential that these words of motivation given by the robot should be at suitable intervals. For example, when encouragement is provided to a disengaged participant, the robot should wait an appropriate period before offering it again, otherwise it could cause annoyance in participants. This interval could depend on a lot of factors, including the reaction of the users to previous prompts, the type of song being played and overall mood of the interaction. Research is needed to determine a relevant interval, but in the current scope, the encouragement prompts are given at the discretion of the human operator, and the interval could typically range from 45-90 seconds.

In the future, automation could be achieved through the use of unidirectional microphones or signal processing techniques which isolate individual voices from background noise. This would allow for more accurate analysis of participant engagement by the robot, without the involvement of a human operator.

3.2.3. Robot Engagement: Pose Estimation

Detection: As mentioned earlier, during the course of the activity, the robot can provide the users with two prompts: either to clap along to the music, or to raise their hands. These prompts are not automated within the robot, but initiated by an activity coordinator (the researcher) in a WoZ technique, who assesses the situation and determines the appropriate prompt to give.

Various indicators can be used to evaluate the apt moment for delivering a prompt. It is important to strike a balance: prompts should not be too frequent, as that might frustrate the user, nor should they be too far apart as the interest towards the robot might wear off. In general, if the operator notices that participants have not been engaged with a robot for a while, such as not making eye contact with the robot, then a prompt can be provided. This interval could differ based on the situation, and could vary between 60-90 seconds before considering another prompt. However, more research is required to determine the optimal duration between consecutive prompts. The type of song playing might also influence the nature of prompt. For instance, if the song is upbeat and rhythmic, users can be asked to clap with the beat. On the other hand, if the pace of the song slow, participants can be asked to raise their hands and sway to the music. Another important factor is whether to give higher preference to activity engagement or robot engagement. In this context, if a user is not singing along, we prioritize improving their activity engagement and focus on providing encouragement to participate in the activity. Activity engagement takes precedence in this case, because without singing along, the effectiveness of the robot prompts may decrease - we do not expect users to sway to the music if they are not singing in the first place.

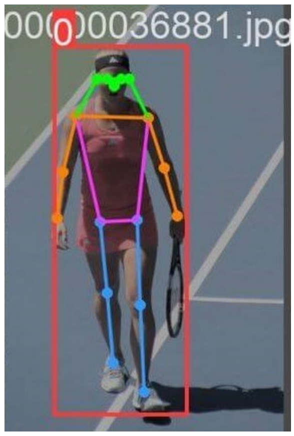
To determine whether the users are responding to the robot, we use pose estimation to assess whether they are performing the said activity, namely clapping along or raising their hands. For this, we make use of the pre-trained YOLOv8 pose model by Ultralytics[7]. The model has been trained on the COCO dataset [92], specifically COCO128-pose[8], which is a large scale dataset used for tasks like object detection, pose estimation and image captioning. The YOLOv8 pose model takes an image as input and returns the X and Y coordinates of 17 keypoints of each detected human figure, namely the nose, both eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles. It also returns the X and Y coordinates of the boxes bounding the humans in the images. In short, each human figure in the image has one bounding box with X and Y coordinates. Each bounding box consists of X and Y coordinates of the 17 keypoints of the individual's body. A few examples of the keypoints and bounding boxes are shown in Figure 3.3a and Figure 3.3b.

Here are the steps to detecting whether the participants are responding to the robot:

- The pre-trained model is loaded into the Python code.

[7] Ultralytics YOLOv8 pose model
[8] COCO pose dataset

(a) Keypoints of human figures



(b) Keypoints and bounding boxes of human figures

Figure 3.3: Output of pose estimation using the COCO pose dataset (Source: Ultralytics COCO pose)

- The robot prompts users to either clap or raise their hands.
- Subsequently, the webcam records an 8-second video of the participants at a rate of 30 frames per second.
- Each frame is then analyzed in real time to determine the X and Y coordinates of various body parts of the each user using the YOLOv8 model, including the left and right wrists and elbows.
- The X and Y coordinates of the bounding box for each individual is also stored.
- The keypoint coordinates are processed to analyze whether the user is responding to the robot or not:
  - To detect if the users are clapping, the euclidean distance between the left and right wrist is calculated and monitored using the respective X and Y coordinates. If this distance consistently varies, alternating between increasing and decreasing, we infer that the individual is clapping.
  - To identify if the person is raising their hands, the Y coordinates of the elbows and the wrists are compared. If the wrist is higher than the elbow in 60% of the captured frames (within a duration of 8 seconds after the prompt), the individual is considered to be responding to the robot.
- As this activity involves multiple participants, we need to identify the individuals that are not engaged. To achieve this, we use the X and Y coordinates of bounding boxes, sorting them in ascending order. This arrangement provides a sequence of keypoints, corresponding to users from left to right based on their seating positions. Next, we utilize the participants' names which have been input into the code according to their seating arrangement and associate each name with the respective bounding box. Now if any user becomes disengaged (as calculated from keypoints), we can identify them by name.

Re-engagement Strategy: Once the disengaged participants have been identified, namely those participants who do not respond to the robot's prompts, the robot addresses them by their name and encourages them to join the group. Along with this, the robot uses calming gestures to reassure and soothe individuals in case they are anxious to join the activity [4]. This serves to create a personalized and inclusive atmosphere, where all participants are supported and motivated to engage in the activity.

It is possible that an individual is actively engaged, but does not perform the actions that the robot prompts. In this case too, the robot would address the individual and provide encouragement. It would just serve as additional motivation for the user. By acknowledging and encouraging all participants, regardless of their response to the prompts, the robot also creates a positive environment conducive to enjoyment for all involved.

3.3. Pepper Robot

The implementation of the project is done using the Pepper Robot, manufactured by SoftBank Robotics [9]. (see Figure 3.4). In this section, we discuss about Pepper in more detail, along with the motivation for choosing it, followed by the frameworks utilized for coding the same.

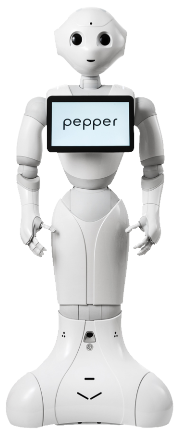


Figure 3.4: Pepper Robot (Source: Aldebaran Pepper Robot)

3.3.1. Technical Specifications

The Pepper robot[10] is 1.2 meters tall, and weighs 28 kilograms (see Figure 3.5). It has an attached 10.1 inch touchscreen display, and is incorporated with touch sensors, cameras and microphones to facilitate multimodal interactions. The two High Defintion (HD) cameras are mounted on its forehead and in the mouth, and all four microphones are on the head of the robot. The head and hands have touch sensors, which enables customization of the robot reaction when touched. It can interact in 15 different languages, including Dutch and English. Additionally, the robot has 20 degrees of freedom, which allow natural and human like movement of the head, shoulders, elbows, wrists, fingers, hip, knees and base (see Figure 3.6). The robot can walk around at a speed of 3 kilometers per hour independently, and navigates its path with the help of laser, sonar, bumper and gyro sensors attached on the legs. A lithium ion battery is used, and can run around 12 hours on battery without having to recharge it.

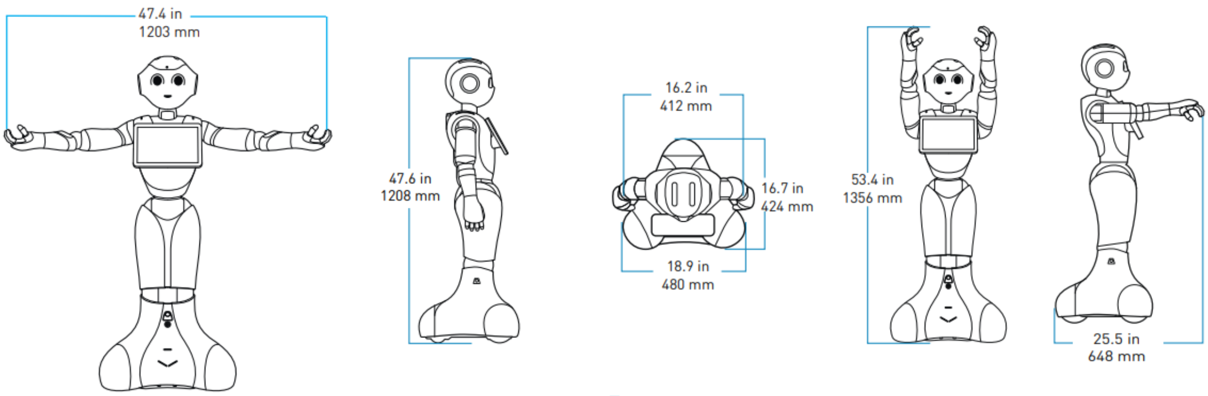


Figure 3.5: Pepper Size (Source: Pepper Technical Specifications: Physical Information)

[9] Softbank Robotics: Pepper
[10] Technical Specifications: Pepper

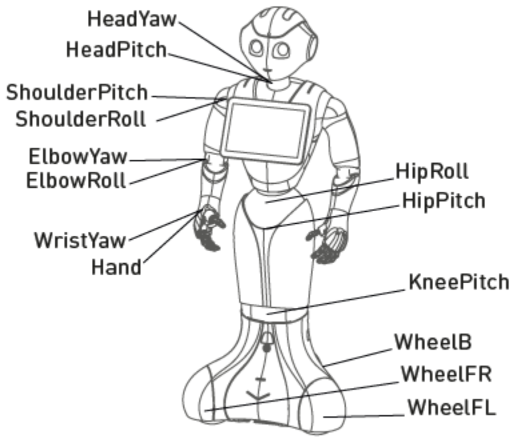


Figure 3.6: Pepper movements (Source: Pepper Technical Specifications: Motion)

3.3.2. Why Pepper?

There were several compelling reasons to use the Pepper robot in this study. The primary reason Pepper was chosen was its height, which contributes to its human-like appearance and aids interaction with users [93][94]. Standing at 1.2 meters tall, it is comparable to the height of an individual sitting on a chair [95]. Its stature distinguishes it from other shorter robots, which have limited visibility and might go unnoticed. This makes Pepper easier to spot and interact with, and is particularly beneficial for older adults in a group environment. This optimal height also enhances the ability to detect people, interpret emotions, understand poses and establish eye contact, thus making it more anthropomorphic.

Moreover, Pepper is designed to be a friendly and engaging social companion and assistant. It is incorporated with basic emotional intelligence and people tracking abilities, which helps it to connect with people. Since the robot is a SAR in the scope of the thesis, Pepper is the evident choice. It is also available in multiple languages, making it possible to adapt to the local language, which in this case is Dutch.

The robot also comes attached with a tablet, which can be used to display relevant information, images and even play videos. An improved feature of the project could allow users to select and play their desired song by simply tapping on the tablet screen. This project could be further enhanced with games that can be played using the tablet.

Finally, being a widely used robot in HRI, there are multiple frameworks to program and customize its behaviour. Numerous researchers [42][59][70] have leveraged these tools and frameworks to explore and create their version of a humanoid robot. A number of these frameworks are explored, further detailed in the next section.

3.4. Implementation Frameworks

There are multiple ways and tools available to program the Pepper robot. The pros and cons of each framework were assessed with regard to the current study, and the selection of the framework was done accordingly.

3.4.1. Pepper SDK

The Pepper Software Development Kit (SDK), also called the NAOqi SDK, is the official framework developed by Softbank Robotics for programming the robot. It provides support for many popular languages, including Python, C++ and Java. It consists of libraries and toolkits through which developers can program the Operating System (OS) on the robot, NAOqi, which is the middle layer, connecting the software and the hardware of the robot.

The NAOqi SDK is one of the most flexible ways of working with and customizing the robot actions, since it works directly with the OS. An extensive and detailed manual[11] is also openly available to describe available

[11] NAOqi SDK Development Guide

APIs, which serves as a valuable reference to understand the range of options and capabilities of the robot. It is noteworthy that the Pepper and the Nao robot use similar functions and APIs, making it a potentially seamless task to transition from one robot to the other. While the SDK offers considerable flexibility, the SDK is challenging to work with and requires in-depth knowledge of the NAOqi OS to effectively write code. Although the manual is openly available, it's structure is confusing, and there are not enough resources to understand the syntax of API usage and practical aspects of code implementation. A simple task too, makes use of multiple APIs and functions, which can get quite challenging to implement. Finally, the Python SDK works only on an older version of Python, specifically Python 2.7. This makes it infeasible to use, especially when the program requires packages exclusive to Python 3, such as several machine learning and deep learning libraries like TensorFlow and Keras. Due to these limitations, the Pepper SDK is not used for implementation.

3.4.2. Choregraphe

Choregraphe is the graphical user interface developed by SoftBank Robotics to create applications for humanoid robots, including Pepper and Nao. Available on Linux, Windows and OS, this application allows users to model robot behaviour by using pre-defined actions and behaviours and combining them in a sequential fashion, as depicted in Figure 3.7. Choregraphe enables users to drag and drop options from the menu to create their own flow, thus making the process very simple and intuitive even for those without programming expertise.

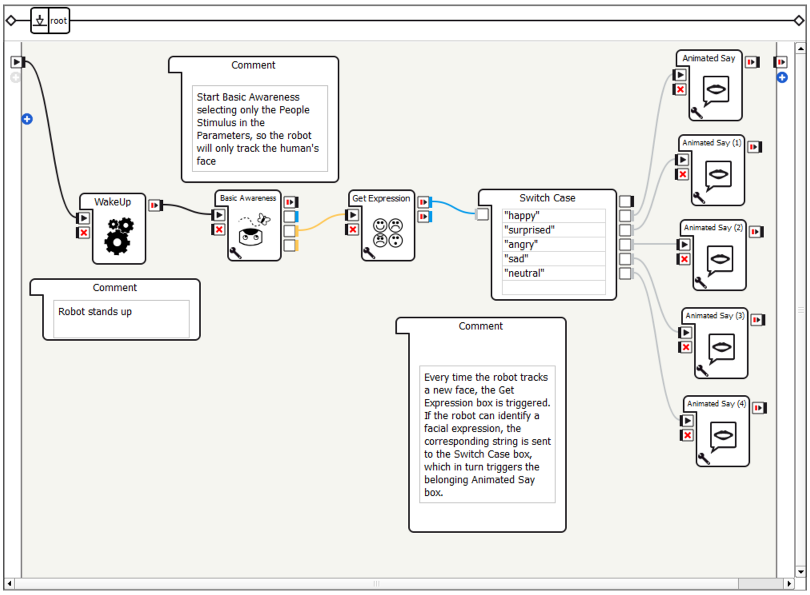


Figure 3.7: Choregraphe (Source: Face Characteristics: Get Expression)

It has a wide variety of features, including Face Emotion Recognition (FER), recognizing voice inputs and motion and animations. All these operations can be performed in just a few clicks, making it incredibly efficient to use. The flow of FER is shown in Figure 3.7. Apart from being easy to understand, the tool provides a virtual or simulated robot on which most behaviours and actions can be tested before deploying it on an actual robot. Similar to the Pepper SDK, it has a step-by-step installation and development guide[12], which is useful for understanding available functionalities and getting started with customizing behaviours for humanoid robots. Additionally, users can also integrate custom Python code, thus enhancing the versatility. Given these array of features, this tool is favourable for programming simple behaviours, preferably those which do not require computation or artificial intelligence algorithms. However, it becomes difficult to integrate extensive Python code, especially where the user is required to load models, process and analyze video data. The proposed flow requires precisely this for engagement detection, which makes it unviable to use in the current context. Additionally, inbuilt capabilities requiring the use of sensors, like FER and voice recognition might not function flawlessly, and users might require more advanced models to ensure optimal performance. Finally, it must be noted that the tool requires a license key to be purchased after the end of a one month trial period.

[12] Choregraphe Development Guide

3.4.3. WoZ4U

The WoZ4U framework [70] was developed for human robot interactions and experiments using the Wizard of Oz technique. As explained in Subsection 2.3.4, this is a technique used in the initial stages of the application to gather preliminary feedback. It involves a human operating the robot from behind the scenes, while the participants interact with the robot under the impression that it is working autonomously.

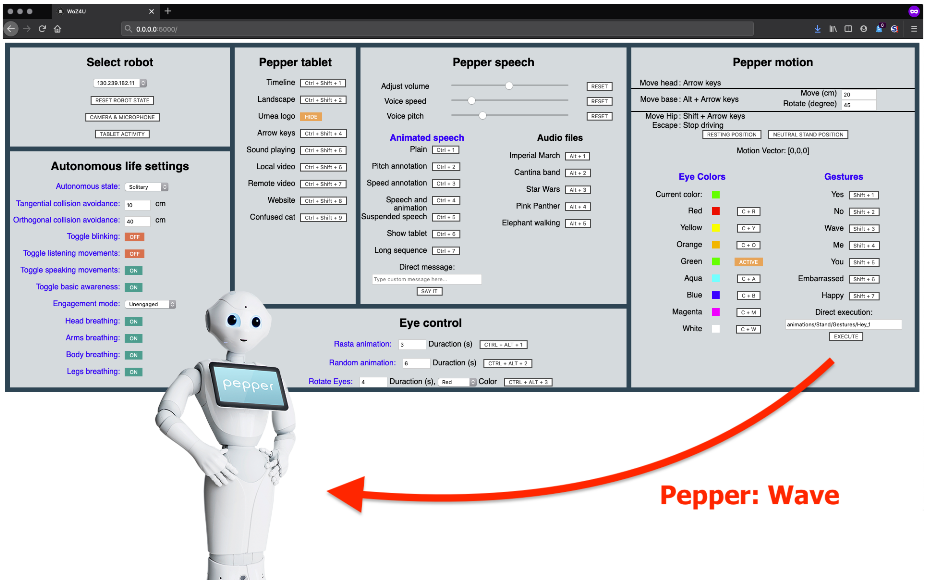


Figure 3.8: Choregraphe (Source: WoZ Interface [70])

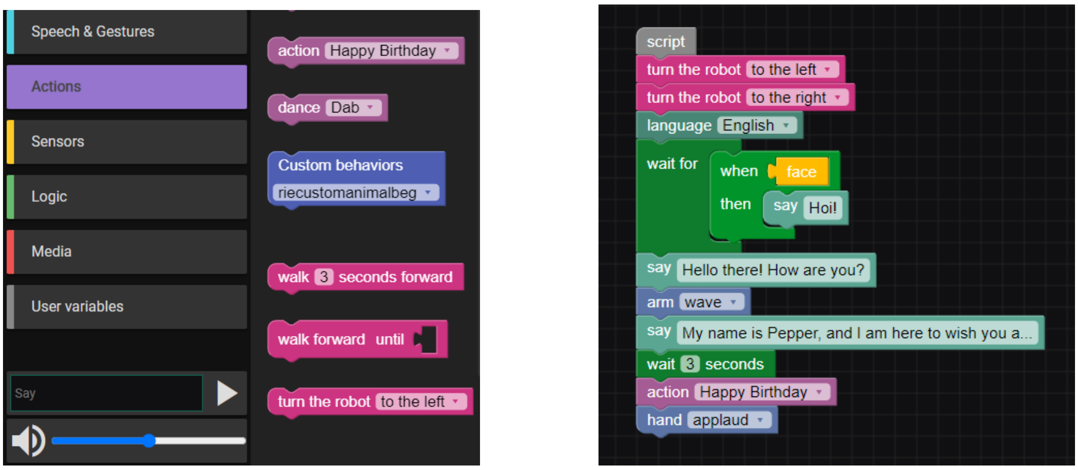
The Woz4U framework provides a web interface which when connected to the Pepper robot, offers a range of control options, including speech, motion, gestures and various autonomous life settings. The interface is shown in Figure 3.8, and is customizable. For instance, if we would like the robot to say something, the text can be typed into the designated space provided under 'Pepper Speech' in real time. As soon as the command is given, Pepper promptly performs the said action. The framework is very intuitive and easy to use, giving researchers the capability to control the robot behaviour in real-time. However, it is not possible to perform more complex computations with this and integrate it with Python code. As a result, relying solely on it for executing custom operations becomes unfeasible. Moreover, it requires installing various packages on the local system to make it functional, resulting in an added overhead.

3.4.4. Interactive Robotics: Robots in de Klas GUI

The Robots in de Klas Graphical User Interface (GUI)[13] is a cloud-based software platform, powered by Interactive Robotics[14]. Mainly developed for purposes of education and learning, it provides a foundation for research in the field of social robots. Alongside a user friendly interface to seamlessly connect to a robot, it uses pre-defined and customizable behaviours and actions, as shown in Figure 3.9a, that can be implemented on a robot. Similar to Choregraphe, the actions can be dragged and dropped to create the desired flow (see Figure 3.9b). It has various options of robot actions, sensors, logic, gestures and speech, which can be included in the flow as needed. Not only can the created flow be executed on physical robots such as Pepper and Nao, but the platform also provides a virtual robot for testing purposes.

In contrast to Choregraphe, the Interactive Robotics GUI does not require installation of any application, or a license key. It operates as a website, accessible from any device with an internet connection, thus providing easy accessibility. As mentioned, it provides a plethora of options for controlling the robot, ranging from miniscule tasks like changing the eye colour to physically maneuvering the robot. Additionally, it offers real-time control of the robot with a WoZ technique. For instance, one can input the text they want the robot to say directly, separate from the predefined flow. Upon submitting the command, the robot promptly repeats the provided text. It is a user-friendly platform, and has multiple sample applications in the library, which can be used as a reference or a base to start development from. Having various inbuilt languages, the

[13] Robots in de Klas
[14] Interactive Robotics



(a) Pre-defined actions and behaviours

(b) User defined flow using pre-defined actions

Figure 3.9: Robots in de Klas GUI

platform also provides the feature to change the robot language to suit the participants. Moreover, it has various robot modes, including *talkative* and *receptive*. The *talkative* robot is chatty, and attempts to keep the user engaged even outside the flow of the program, whereas the *receptive* robot focuses more on listening to the participants and does not talk outside of the program flow. Having said this, it is important to note that the GUI alone might not be sufficient for algorithms involving machine learning models to process and analyze data. Since a more comprehensive approach is required for engagement detection and pose estimation, we do not use the tool alone, but in conjunction with another framework which provides the opportunity to easily write custom code to integrate into the robot.

3.4.5. Interactive Robotics: Advanced Programming

Interactive Robotics also has an advanced programming version in addition to the GUI tool described in the previous section. This allows users to connect to robots and program them via Python or JavaScript code. The manual for the same can be obtained by sending an email to the team. For communication with the robot, a wamp connection is established using the autobahn package along with the 'realm' of the robot. A 'realm' is defined as a closed environment in which the robot resides. In other words, it is the identification of the robot that the users want to connect to. The 'realm' of the robot the user wants to connect to can be obtained from the Robots in de Klas GUI portal.

The framework manual serves as extensive guide to programming the robot, with instructions on installation of packages, creating the connection, issuing commands and monitoring sensors, along with sample code for reference. Offering the framework in two languages increases its accessibility and usability to a wider range of users. Additionally, writing code in Python allows developers to perform computations and write algorithms, thus enhancing flexibility. Since the realm establishes a link between the code and a robot on the Robots in de Klas GUI, the code can also be tested on their portal's virtual robot before it is deployed on a physical one. However, the learning curve with this framework is quite steep, necessitating understanding of its commands to fully utilize the multiple features it provides. It is also not possible to provide commands in real time, without including these in the code prior to execution. Finally, the installation of the *autobahn* package with the optional dependencies *twisted* and *serialization* gave errors in a conda environment of Python 3.8.

3.4.6. Social Interaction Cloud (SIC)

The Social Interaction Cloud (SIC) [15] has been developed by Koen Hindriks and his team at VU Amsterdam. The framework, still under development, provides a way to program the robot using Python. The connection is established through a Redis server, which needs to be up and running throughout the duration of the flow. The framework can be installed on the local system via git, and after the installation of a few libraries, it is ready to be used.

[15] SIC

The framework can be installed and used on Windows, Linux and MacOS, thereby enhancing versatility and usability across users with different operating systems. Not only does it provide a step-by-step guide into installation of the framework, but also video explanations for better understanding. SIC components, which are the core building blocks, are used to combine different services and data processing tools into the desired program. For example, code for frequently used services like face recognition are written as an SIC component, which can be included into any code, thereby eliminating the need of re-writing the same code and re-inventing the wheel. SIC components for face recognition, displaying an image from the robot camera, person tracking and transcribing audio are few of the many services inbuilt with the framework. In addition, users can also create their own SIC components to use as building blocks in the code.

Hence, it provides a flexible way for the developer to use their services, and a platform to communicate with the robot. Moreover, the functions of robot speech, gestures and poses are easy and intuitive to implement. What makes the platform powerful is the level of abstraction it offers. For instance, the programmer simply needs the IP address of the robot to establish a connection, without concerning themselves about intricate protocols and underlying mechanism. Once the robot is connected, the framework promptly initiates required services like *TextToSpeech*, which are then accessible with a single command. It also has various programs for demonstration purposes, which can serve as a valuable reference. Since the implementation is done using Python, custom code to use machine learning models and perform computations is easy to integrate in the system. Although the system boasts simplicity for basic operations, it's status as a work in progress may occasionally result in incompleteness in the code and documentation. The documentation lacks comprehensive details of the available options and features, making it difficult to completely utilize the offered functions. As a result, one must either refer to sample code to perform operations, or delve into the back-end framework code. Sometimes, the written code may fail to execute as expected, and pinpointing the exact reason can be challenging. For example, the *TextToSpeech* service takes language as an input, however, attempting to change the language during the flow does not work. The provided Basic Awareness SIC component too failed to give desired results, causing the Pepper robot to unexpectedly turn to the extreme left without a clear understanding of the underlying issue. Moreover, real-time commands cannot be directly issued to the robot using SIC. Instead, all commands must be pre-written as code and executed accordingly. Finally, implementing this on a Windows Operating System did not work with the given instructions, requiring the installation of Linux to use it.

3.4.7. Discussion

We summarized the key features of 6 different frameworks that can be used to develop a robot, along with analyzing their advantages and disadvantages. While GUI tools such as Choreographe, Interactive Robotics: Robots in de Klas and WoZ4u are easy to use, they tend to be oversimplified for the current context, making it difficult to write custom code. The code-based tools like Pepper SDK, Interactive Robotics: Advanced Programming and SIC provide a lot of flexibility with the ability to integrate custom code. However, with limited documentation and samples, the learning curve could be considered steep.

Keeping in mind the pros and cons, and the requirements of the thesis, we choose to work with a combination of two frameworks: the **SIC framework** discussed in Subsection 3.4.6 and the **Robots in de Klas GUI** described in Subsection 3.4.4. Both of these are connected to the robot, and executed as required. The bulk of the programming is done using the SIC framework which oversees most tasks including welcoming participants, presenting song choices, and managing webcam input. This input is then processed using a pre-trained model for pose estimation (see Subsection 3.2.3), enabling the system to assess user engagement and accordingly evaluate the appropriate robot response. The Robots in De Klas GUI is only used to control the robot with real time commands for responding to activity engagement as SIC does not support this. If the human operator, who serves as the Wizard of Oz, notices that a participant is not singing along, they use the GUI interface to type a message addressing the participant and encouraging them, which the robot promptly verbalizes (see Subsection 3.2.2). Moreover, the GUI is used to instruct the robot in real time to execute gestures such as dancing, applauding and raising hands when deemed appropriate by the WoZ.

This is how the two frameworks work in tandem to provide a seamless flow. While the SIC carries the brunt of the work, the GUI is primarily used to issue prompt and real-time commands to the robot.

4 Experimental Method

The implemented prototype is evaluated with experiments to gain feedback and answer the research questions posed. For these experiments, groups of older adults are invited to the Insyght Lab at TU Delft. They interact with the implemented robot, which guides them through one session of a sing-along activity. The participants are then asked for feedback about the robot, its behaviour and actions, along with the efficacy of the encouragement it offers. Inquiries about the activity are made, to judge whether it has potential to improve the moods of participants and uplift them. Their social engagement is analyzed, which pertains to the how participants of a group interact with each other. Deeper insights into the same are gained with the analysis of the recorded video sessions.

With these findings, we can address our research questions. The first research questions revolves around perception and response of older adults to the designed robot. Questionnaire responses and the video analyses provide an understanding about their responsiveness towards the robot and their acceptance into the group of humans. We study the effectiveness of the interaction, the impact of the robot on the participants, and their general attitude towards it.

The second research question pertains to the social engagement and bonds formed within the group of participants. We observe the types of interactions they have with each other during the session, the group dynamics and how they make decisions together while selecting a song. Their behaviour during the session is monitored to analyze aspects of their interactions, such as: Are they an active member of conversations? Are they encouraging each other to sing? With this, we try to understand whether such an activity can improve their social engagement and whether it can help them become closer to their peers.

4.1. Pilot Study

Before conducting the experiment with participants aged more than 60, an initial pilot was performed with students of TU Delft, for initial feedback. Two sessions were conducted, with group sizes of 5 and 2 respectively. Although these participants do not represent the target audience, these experiments provided valuable feedback for system improvement.

At first, the robot focused on providing encouragement only to the users not singing along. However, during the pilot it was noticed that the robot must also motivate those who are actively participating, and consistently keep interacting with them, through verbal motivation and non-verbal movements. It was also observed that the robot could not be heard very clearly over the music, requiring a solution that lowers the volume of the music while the robot speaks. Suggestions about the choice of music and songs were also received and incorporated.

4.2. Participants

The target group intended for this experiment is individuals with dementia, or cognitively impaired persons. However, to gather initial feedback, we decide to work with older adults, those above the age of 60. Participant recruitment was done through choirs in the Delft area, who were approached, and emailed with a poster consisting of the details of the experiment. Those who were interested in participating responded to the advertisement, after which a student assistant organized them into groups according to their availability and confirmed their slot for the experiment.

A total of 8 participants (apart from the initial pilot study) participated in the experiment. This led to 2 groups of 4 members each. There were 7 female and 1 male participants, all above the age of 60. The experiment was conducted in the Insyght Lab of Building 28 in TU Delft, where the participants were asked to come in. Each experiment session lasted around 45-60 minutes, including all the tasks from starting with welcoming the participants and explaining the experiment procedure. This was followed by signing of informed consent forms, performing the actual experiment and the completion of questionnaires and surveys.

4.3. Research Design

The experiment is designed to answer two main research questions outlined in Section 1.5. In this section, we discuss the experimental design which serves as a roadmap for systematically addressing the research questions while maintaining methodological rigor. The observed variables are outlined, along with the measures used to evaluate and address each of the research questions.

RQ1: How do older adults perceive a robot which moderates a sing-along activity, with real-time engagement detection and re-engagement strategies?

Purpose of Enquiry: The research objective is 'descriptive', as we attempt to *decribe* the characteristics, behaviour and attitude towards a phenomenon, which in this case is the robot. We do not manipulate or change the value of variables, instead just monitor the value of a particular variable, called the dependent or observed variable. In this case, the observed variable is the users' perception of the robot they interacted with. Since there is no independent variable(s) whose value we change, there is no *control* and *experimental* scenario; instead we work with only one condition, observe it's output on the dependent variable and gather relevant data through questionnaires. Though descriptive research cannot be used to establish causal relationships, it can help in generating hypotheses for future research.
Observed Variables: The users' perception of the robot is the variable to be observed.
Measures to evaluate: The Robotic Social Attributes Scale (RoSAS) questionnaire [84] was used to evaluate the users' impression of the robot. A subjective questionnaire with open ended questions was also used, granting participants the freedom to express their thoughts and experiences without constraining them to a chosen set of options. Along with this, a qualitative study was performed to assess behavioural measures during the trial and the activity engagement of users. This is described in further detail in Section 4.4.

RQ2: Can participation in a sing-along session moderated by a robot with a group of older adults enhance interaction with peers?

Purpose of Enquiry: The research objective can be considered as a combination of exploratory and descriptive research. Since utilization of group HRI to promote meaningful interactions is a niche topic, and is not widely researched, we aim to gather some insight into the area. This serves to explore the field and gain a deeper understanding of the dynamics involved, thereby paving the way for innovative approaches for future research. Additionally, the research question falls under the umbrella of descriptive research as we try to *describe* the effects of a group HRI on social interaction within the group. Similar to the first RQ, there is no *control* and *experimental* scenario; we work only with one condition: interaction as a group.
Observed Variables: The feeling of social connect with the other members in the group is the observed variable in this case. This can also be thought of as a sense of community within the group, or interpersonal bonding. It is expected to be the result of interacting in a group. However, since this is not a comparative study, we don't try to establish a causal relation with another variable, instead we gain insights and try to understand the social cohesion given a group interaction.
Measures to evaluate: A subjective questionnaire was used to gather personalized and detailed feedback. Alongside this, the same qualitative analysis as RQ1 was performed to gather social and subtle cues that could be indicators of social engagement. This is described in more detail in Section 4.4.

4.4. Measurements

Three different measurement techniques are employed to address and answer the research questions in a methodical and systematic manner. Not only do multiple techniques offer a comprehensive understanding and thoroughness, but also an enhancement in the reliability of the results. The measurement techniques are outlined below.

1. **Qualitative Analysis using video data:** for RQ1 and RQ2.

To assess activity engagement, social engagement and attitude towards the robot, the entire experiment was recorded. This analysis divided each session into 4 main phases:

(a) Introduction Phase: The robot introduces itself and explains the details of the session, along with what the participants are expected to do.
During this phase, we assess whether the participant is attentive to the robot while it gives instructions, or whether they are distracted. This can be evaluated by the object the participant is looking it. If they are not looking at the robot, it might indicate disinterest. The videos are used to learn the reason (if any) for the distraction. Second, we examine whether participants are interacting with each other. This interaction could be verbal, in which case they would be talking over the robot's voice, or non-verbal, by making eye contact with each other. This behaviour can give us insights into their focus and involvement in listening to the robot. These interactions can be analyzed to potentially deduce participants' attitudes towards the robot and the social dynamics within the group. The mood of the participant through facial expressions and body language is also evaluated. These factors can provide an understanding of the general attitude of the participants towards the robot in the beginning of the experiment.

(b) Song Selection Phase: The robot guides users to selecting a song which they would like to sing as a group.
Similar to the previous step, we analyze the gaze direction of the participants to understand their focus during this phase. The objective is to observe the dynamics within the group when they are asked to make a decision. We look at whether participants are interacting with others and actively contributing to making a decision about the song, or whether they are passively listening. If they are passive, we assess whether other members of the group make an effort to involve them. The mood of the participant is also examined during this phase.

(c) Singing Phase: The main phase of the session, where users sing along to the chosen song, and the robot encourages them.
Similar to the previous phases, we analyze the object the participants are looking at, their mood and whether they are interacting with each other. We observe whether participants make eye contact with each other, encourage each other to sing and talk to each other. This could provide insights into their social engagement and participation in the group. We also evaluate their robot engagement: Are they responding to the robot prompts? Are they looking at the robot? How do they react when the robot offers encourages? Their activity engagement is also estimated by the amount of singing. If they sing nearly 100% of the songs, it indicates a high activity engagement and vice versa. This could be an indicator of their enjoyment of the activity and its potential to improve mood.

(d) Final Phase: The robot thanks users for their participation and asks for feedback.
Apart from analyzing the object of focus for each participant, their level of interaction with other participants, and their mood, this phase focuses on the feedback the participant gives to the robot. Depending on whether or not they provide feedback to the robot, we can judge their attentiveness towards it. The feedback itself provides insights about their experience and highlights areas for improvement.

Upon completion of the trial, each video underwent examination and scoring on the basis of a fixed predefined protocol for each of these phases. This protocol is partly derived from [43] and is attached in Appendix B. This analysis allowed for a deeper understanding of participant behaviour, as compared to relying solely on questionnaires. Unlike surveys, which may overlook subtle nuances, human observers conducting the video analysis were able to notice intricate behavioral patterns, subtle interactions, and non-verbal cues that contribute to a more comprehensive understanding of participants' engagement and attitudes towards the robot. It provides richer insights into the human-human and human-robot interaction, a particularly valuable asset given the exploratory and descriptive nature of this study. Thus, in addition to the questionnaires, this qualitative study enables a thorough exploration of the context while picking up on clues that questionnaires could miss.

2. **RoSAS questionnaire:** for RQ1.
The Robotic Social Attributes Scale (RoSAS) [84] is widely used in the context of HRI, and serves as a means to assess the attributes that play a role in human perception of robots. The scale was used to address and gain insight into of RQ1, specifically, the acceptance and attitude of users towards the robot.

Derived from the Godspeed questionnaire [85], this consists of 18 items, which can be categorized into 3 factors:

(a) Warmth: Happy, Feeling, Social, Organic, Compassionate, Emotional

(b) Competence: Capable, Responsive, Interactive, Reliable, Competent, Knowledgable

(c) Discomfort: Scary, Strange, Awkward, Dangerous, Awful, Aggressive

The participants were asked to rate these 18 items on a Likert Scale ranging from 1 to 9. A rating of 1 indicated that the factor definitely had no association with the robot, while a rating of 9 meant that the factor was definitely associated with the robot. The answers were later compiled to get a comprehensive understanding of the perception of users. As mentioned, since this is a descriptive study, the results of the questionnaire are not compared to a control condition. Instead, they are used to gather preliminary feedback and a general understanding of participants' attitudes towards the robot. The questionnaire is attached in Appendix C.

3. **Subjective feedback using an open ended questionnaire:** for RQ1 and RQ2.
This questionnaire is curated with a set of questions that ask participants to rate certain aspects of the session on a scale of 1 to 5. Each question is accompanied with an open-ended section where participants can elaborate, and provide explanations for their rating. This does not restrict the respondents to select from predefined options, but allows them to provide free form responses in their own words. The survey addresses the following areas:

(a) Overall Experience

(b) Social Experience

(c) Role of the robot and interaction

(d) Other Feedback

Participants are encouraged to express their opinions and experiences in detail, to gather data which can aid future designs. The questionnaire is attached in Appendix D.

4.5. Procedure

The ethics approval for the experiments was obtained from the TU Delft Human Research Ethics Committee, requiring a submission of the Data Management Plan, Informed Consent Form and the Human Research Ethics form to identify potential risks to participants. Since the participants were older adults who might have been more comfortable speaking in Dutch, a student assistant was also available to communicate with them in Dutch. He was also a coordinator during the course of the experiment, and was responsible for coordinating the song selection based on the responses from the group.

Participants were asked to come to the Insyght Lab in building 28 of TU Delft in groups of 3 or 4. Before each session, the setup explained in Section 3.1 was verified and went through a dummy flow to confirm everything was working and in place. Once the participants came in, the procedure was explained, after which they were asked to sign the Informed Consent forms (available in both English and Dutch). Then they were seated on the chairs placed in front of the robot and the television screen, as shown in Figure 3.2. The experiment required two coordinators that took on the role of the wizards and controlled certain aspects from the background. They were seated in the Insyght lab as shown in Figure 4.1.

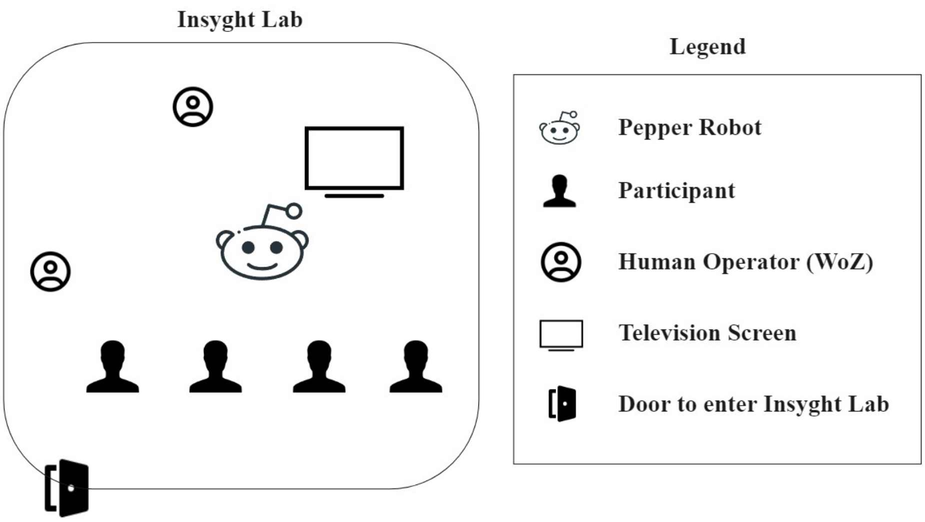


Figure 4.1: Seating Layout

The experiment started with the robot greeting and went through the complete flow, as explained in Section 3.1. The experiment itself took approximately 20-30 minutes. After this, the two questionnaires described in Section 4.4 were given to fill, and we answered any other questions they had. Each group took a total of 45-60 minutes from beginning to end. After the participants left, their names were deleted from the system, and the recording was saved on the device for a qualitative analysis.

5

Results and Discussion

5.1. Analyzing the Questionnaires

5.1.1. RoSAS

In order to analyze the answers to the RoSAS questionnaire visually, the scale was divided into three separate and equally sized bins. Ratings from 1 to 3 indicated that the attribute was not significantly associated with the robot much. Ratings between 4 and 6 indicated that the robot could moderately be associated with that specific attribute. Ratings ranging from 7 to 9 suggest that participants felt a strong association between the attribute and the robot. This allows for a clear analysis of how the robot was perceived by participants in terms of various characteristics.

A visual analysis was performed with the help of a histogram, to see the frequency distribution of the ratings. This could give an insight into whether the distribution is balanced or skewed in a particular direction. These frequency distributions are illustrated in Figure 5.1, Figure 5.2 and Figure 5.3 for the categories of warmth, competence and discomfort respectively.

Warmth:

A strong skewness was found in the *Emotional* characteristic, and all of the responses indicated that the participants did not find the robot emotional. There was a mix of opinions about the robot being *Organic*, *Compassionate*, *Feeling* and *Social* - with a slight skew towards the negative. However, the responses for the robot being *Happy* were balanced. This can be visualized in Figure 5.1.

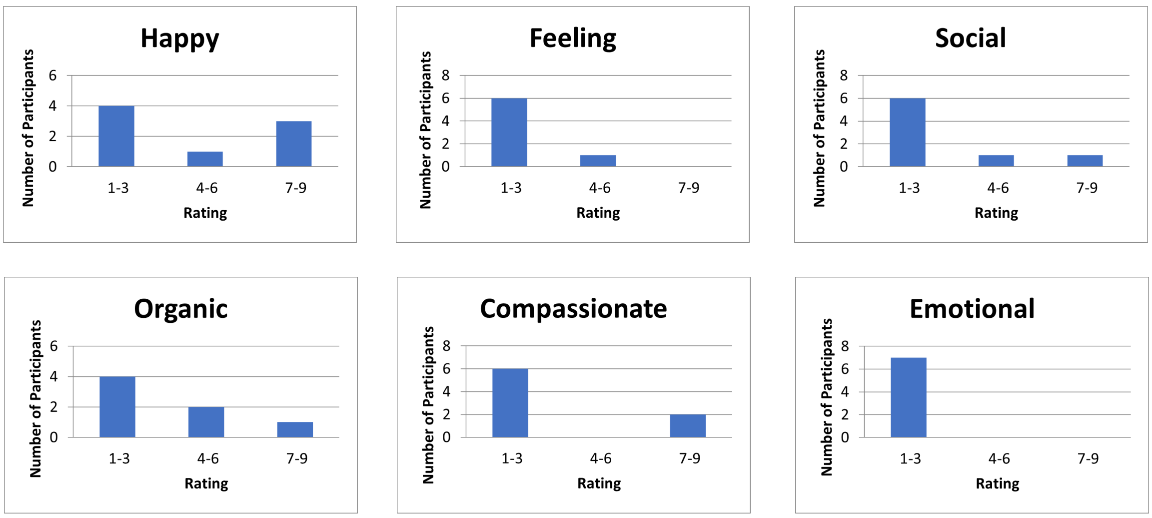


Figure 5.1: Frequency Distributions: Robot Warmth

Competence:

There was a mixed opinion amongst participants about the attributes of competence. A few responses associate a moderate relation with the robot being *Capable*, *Responsive* and *Interactive*, however other participants do not associate the robot with these characteristics. As far as *Reliable*, *Competent* and *Knowledgeable* are concerned, while there is a majority of responses indicating a low score, a handful of responses provide a moderate or high score to the robot in these aspects. All in all, the responses for this category are varied and diverse, but definite conclusions cannot be drawn due to the small sample size.

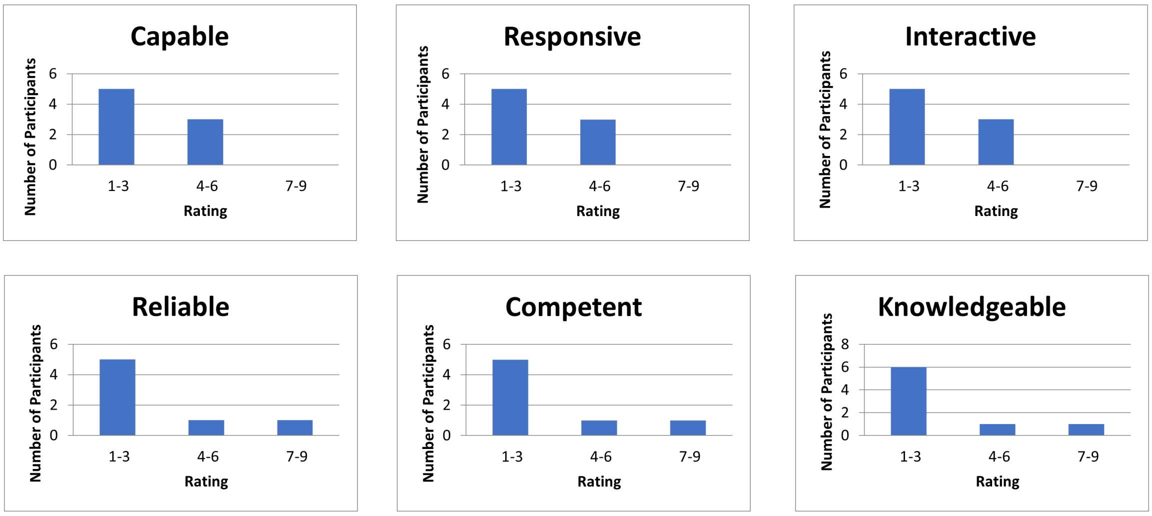


Figure 5.2: Frequency Distributions: Robot Competence

<u>Discomfort:</u>
This division seeks to identify the sentiments of discomfort that participants might have towards the robot. While most of the responses indicated that the robot was not perceived as *Scary*, *Dangerous*, *Awful* or *Aggressive*, a handful of reactions disagree and feel that the robot may possess these characteristics. On the other hand, there were varied perspectives about the robot being *Strange* and *Awkward*; although predominantly being skewed in the lower end of the spectrum.

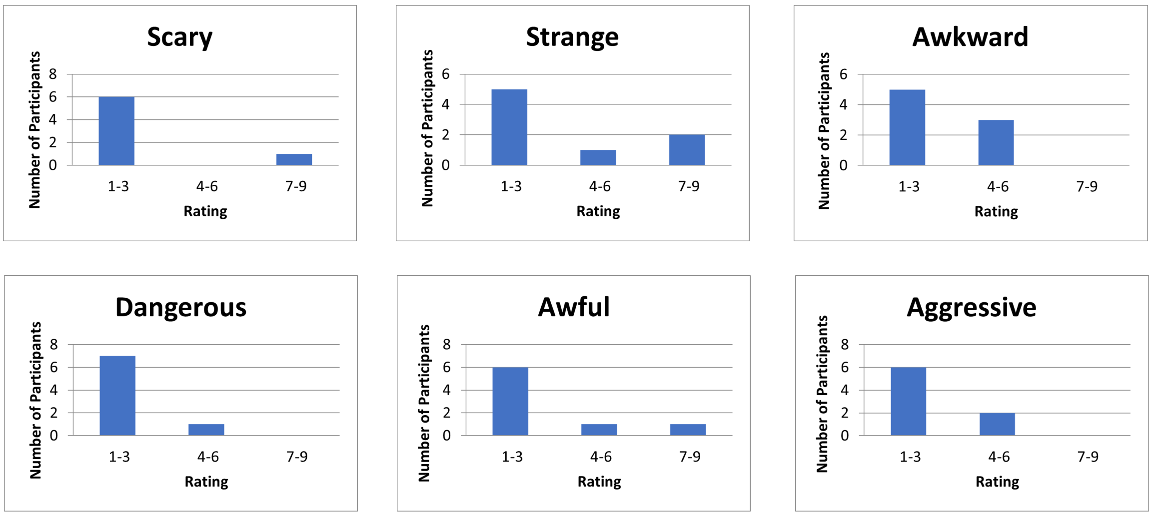


Figure 5.3: Frequency Distributions: Robot Discomfort

5.1.2. Subjective Questionnaire

The subjective questionnaire consisting of aimed to gather feedback on four broad areas, as described below. Each area consisted of questions that could be rated on a Likert scale of 1-5, and each question had an open-ended section where participants could fill in answers in their own words.

<u>Overall Experience:</u>

All participants rated the overall experience as positive, with a median of 4 and a range of 1. The median rating for the activity being able to positively impact mood was 4, with a range of 2. The survey responses are visualized as a frequency distribution to illustrate the general trend of the ratings, as shown in Figure 5.4. The distribution indicates a skew towards higher ratings, suggesting a predominantly positive experience for the participants.

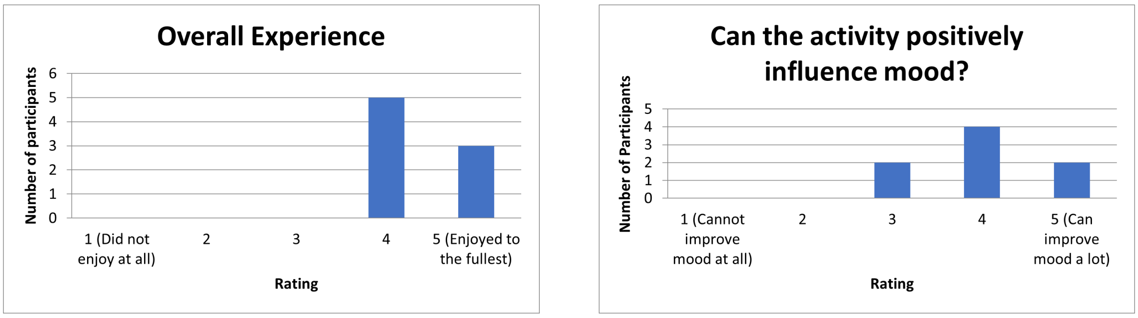


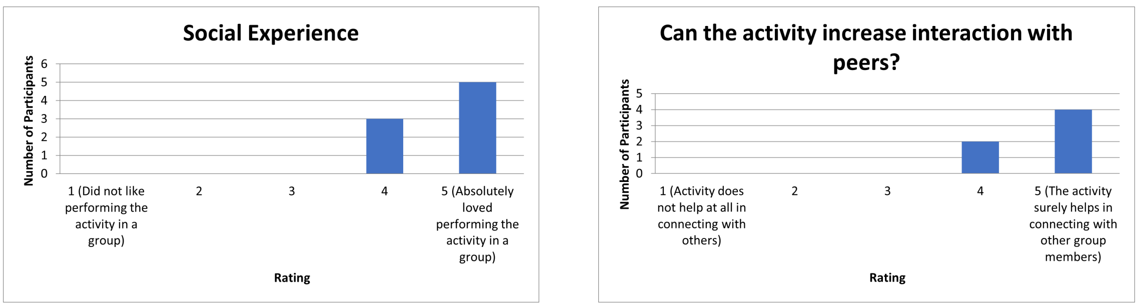
Figure 5.4: Frequency Distributions (Overall Experience)

The feedback provided by participants in the open-ended (qualitative) section of these questions reflect varied perspectives. Overall, all participants enjoyed the activity, expressing that it brought the group together, with one participant stating that the activity *'Brings together people'*. Participants reflect positively on their singing experience, as one participant says *'Singing together is always fun'*. However, participants fail to understand the role of the robot in the session, as they mention *'I really enjoyed singing together. I honestly didn't notice much of the interaction with the robot'*. One participant expresses a tone of frustration towards the robot as they note *'For me, the robot was an obstacle to the pleasant music we sang. The appearance plays a role, plastic, white and weak, and if the robot itself starts singing, there might be more commitment. The voice is also not pleasant'*.

The group also agrees that singing together can positively impact mood, but they do not see the impact of the robot. For example, in the feedback they state that *'I don't think the robot increases mood, but the group does'* and *'Can increase mood, but wants to see more interaction from the robot'*. However, one participant accepts that if deployed outside a lab environment, the robot has the potential of being helpful as they mention *'Do not see value of robot for mood improvement, but if deployed outside lab environment, it might be helpful'*.

Social Experience:

Since the aim of the session was to enhance the connection and interaction between peers, two questions were included to get an insight into whether a robot conducted group activity session can achieve this objective. All participants rated the social experience as positive, with a median rating of 5 and range of 1. All participants also responded that such a group activity has the potential to improve the connection and relations within the group, with a median of 5 and a range of 1, similar to the previous question. The plotted frequency distribution of the ratings is shown in Figure 5.5. The frequency distribution is tipping towards the positive side, with all participants giving a rating of 4 or 5.



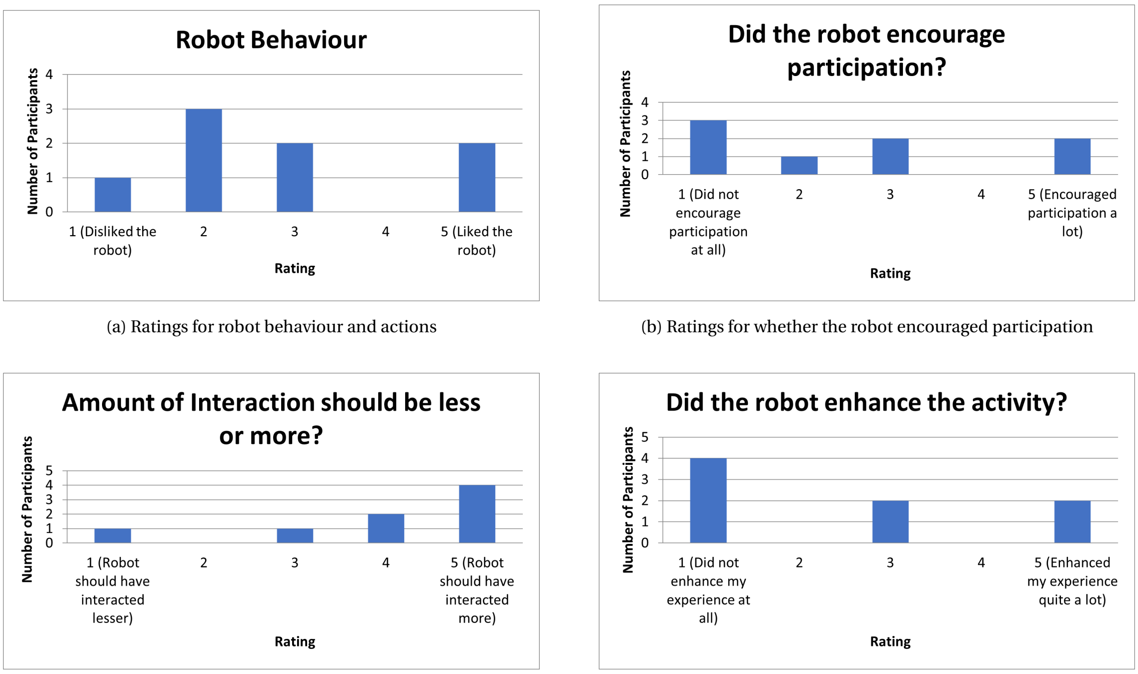
(a) Ratings for the social experience

(b) Ratings for whether the activity can improve connections within the group

Figure 5.5: Frequency Distribution of Scores (Social Experience)

Participants did not know each other before the experiment, and yet found the social experience filled with energy and positive vibes, as one participant mentions *'We didn't know each other, but it was great fun to do this together. There was definitely a positive vibe in the group'*. Others state that it was a *'Cheerful and social get-together'* and *'I loved it'*. Participants also mention that the activity definitely helps connect with others in a group, like most group activities. For instance, they say *'Like many other activities you do with a group, singing together certainly creates bonding'* and *'A group activity always provides connection, but music/singing at any level provides connection'*. If done over multiple sessions, it has the potential to forge lasting connections among individuals, as noted by a participant who says *'Yes (it will help in bonding), but then I have to do a follow-up to it'*.

Role of the Robot:
Various questions were asked to understand what kind of role the robot played and how it was perceived by the participants. The frequency distribution of the responses show an even distribution of ratings across the scale for the questions, as depicted in Figure 5.6.



(a) Ratings for robot behaviour and actions

(b) Ratings for whether the robot encouraged participation

(c) Ratings for whether the amount of interaction was appropriate

(d) Ratings for whether the robot enhanced the activity

Figure 5.6: Frequency Distributions of Scores (Robot Experience)

The analysis revealed that the median score for the likeability of the robot's behavior and actions was 2.5 with a range of 4, suggesting a response hovering around neutrality but high variability in responses. This means that participants had mixed perceptions about the robot's behaviour, neither too positive nor strongly negative. The distribution of survey responses depicts the same, as shown in Figure 5.6a. There was a diverse range of opinions among participants, where some were disappointed and say *'I was curious about the robot but I was very disappointed. more disruptive than adding anything'* and *'The robot is still under development and still needs a lot of work'*. Some had a neutral response, and mention that *'We were singing together and the robot didn't really have a very active role in that'*. One participant liked the way the robot looked and notes that *'The robot moved its arms slightly but did nothing else. at a glance it was beautiful to see'*.

Similarly, a median rating of 2.5 and a range of 4 was observed for the efficacy of the robot's encouragement of participants. As shown in Figure 5.6b, the distribution has high variance, with responses spread across the entire scale. While some participants appreciated the amount of motivation given by the robot, others felt that the robot did not encourage them at all. One note of feedback highlights that the presence of the robot itself made them curious, which fueled their motivation to sing. They say *'Yes (it encouraged me) because the robot made me curious. I see certain possibilities, but the robot would have to be a bit more capable'*. Other participants mention that despite the robot's attempts to verbally encourage them, they were already singing enthusiastically, rendering the robot's efforts ineffective. For example, the participant mentions that

'Did not really encourage. He occasionally shouts something, but we were already singing so enthusiastically that I didn't hear it'. It is important to note that the session took place in a lab environment, where participants attended with the purpose of singing, indicating that they were already highly motivated.

Inquiries about the robot's level of interaction revealed an average response indicating a desire for increased engagement. This was reflected in the median rating of 4.5 and a range of 4, where a score of 1 implied a preference for lesser interaction, 5 meant a preference for greater interaction with the robot and a score of 3 meant the amount of interaction was sufficient. As illustrated in Figure 5.6c, the distribution is skewed towards participants preferring the robot to interact more. Participants observe the robot as *'Passive'*. One participant recognizes the robot's role in guiding them through the activity and states that *'I expected more interaction. it was really the case that the robot helped us through the session'.* Another user mentions that though they liked interaction, more of it would be appreciated, as they say *'More interaction (expected), but good'.*

Participants were also asked whether the presence of the robot enhanced the activity. This received an median score of 2, with a range of 4, which means that the robot might not have had the required impact. Some found the robot impersonal and others said that *'I thought the robot looked funny. but other than that it was mainly singing together that made this a fun experience'.* Another perspective acknowledges that the robot gave them a platform to perform such an activity. The participant mentions that *'(The robot is) funny, but didn't contribute much. It's not easy to sing without a robot. you definitely need some stimulant'.* The distribution of the responses is shown in Figure 5.6d.

Other Feedback:
When asked for any other feedback that participants might have, one participant said that the compliment the robot gave at the end seemed a little random, and it should be given throughout the duration of the experiment, preferably when the music is quiet, so that it can be heard better. They say *'(The robot can) speak more complimentarily at a time when the music is quiet for a moment'.* An alternative proposed was displaying the compliments on the screen mounted on the Pepper robot's chest. One of the participants also suggested making the robot look more attractive. For example, they state *'Other decoration: e.g.: clothes'.* Moreover, the voice of the robot felt awkward. They also felt that personalized feedback for each participant could be incorporated at a later stage; for example, if they sing too loudly or don't sing well.

5.2. Analyzing the Videos

A qualitative analysis was performed on the video recordings to analyze the non-verbal cues and behaviour exhibited by participants. The rubric to assess these behaviours is derived from [43] and has been explained in Section 4.4. The captured videos were annotated and assessed using the rubric in Appendix B. The two recorded sessions were divided into the 4 phases described in Section 4.4, and each aspect was evaluated according to the template. One of the participants was delayed to the session, hence they were not present for the whole duration of the experiment. So we analyze the video data of only 7 out of the 8 participants. In this section, we summarize the **observations** made with reasoning and evidence from the recorded videos.

5.2.1. Robot Engagement and Attitude towards the Robot

As explained in Subsection 3.2.3, we define robot engagement as the extent to which participants actively listen and respond to robot prompts and statements. From the video analysis, we observe that most of the participants actively listen to it. When the robot speaks, it is seen that they stop talking to each other, come closer to the robot and make an effort to listen to it. At the end of the session, they also respond to the robot's question about their experience and their feedback about how the session made them feel.

However, we notice that when the robot prompts participants during the session to clap or raise their hands with the music to detect their engagement, the participants do not perform the said action. However, this might not imply that participants are not engaged with the robot; it could be because the robot was not very audible along with the background music. This is evidenced from the videos where participants lean in to listen to the robot, but the voice is not very clear, hindering their comprehension of the message. This is reinforced by questionnaire responses shown in Table 5.1.

The lack of response to the robot prompts might also stem from the participants already being engrossed in dancing, thus not feeling compelled to clap along or raise their hands. This indicates that their activity engagement was high, but robot engagement was low, as also highlighted in the objective questionnaire responses where participants rate the robot as passive and awkward as seen in Subsection 5.1.1. This evidence

Question	Response
Reflecting on your experience, could you please rate your overall enjoyment of the session, considering both – the robot and the activity?	*'...I couldnt understand it (the robot) well, but the occasional positive comment was nice'*
Did the robot actively encourage participation?	*'Not really. He occasionally shouts something, but we were already singing so enthusiastically that I didn't hear it'*
Do you have any other feedback you would like to give to improve the system?	*'The compliments speak at a moment when the music is silent for a moment'* *'Write compliments on the screen on the robot's chest'*

Table 5.1: Questionnaire Responses: Unclear Robot Speech

is summarized in **Observation 1**.

Observation 1

While participants actively make an effort to listen to the robot's speech, they do not respond to the prompts given by the robot in the duration of the singing session.

The videos also indicate that the attention given to the robot gradually decreased over the course of the session. Participants were initially very curious to see the robot, its behaviour, and its role in the activity. A participant also notes that this curiosity fueled their motivation to sing in the beginning. Initially, the participants are noticeably attentive, keenly listening, and looking at the robot whenever it speaks. However, this changes as the session progresses, with participants becoming more passive in their interactions. This could be attributed to multiple reasons. Firstly, the robot dialog is scripted and there was only a small set of dialogues to choose from, making the robot's speech predictable and repetitive. Second, the strange voice and perceived passiveness of the robot, as noted in the subjective questionnaire responses shown in Table 5.2, could have deterred them from maintaining their focus on it. Finally, the initial intrigue may have diminished as they got accustomed to the robot's presence, resulting in reduced attention towards it. These findings lead us to **Observation 2**.

Question	Response
Reflecting on your experience, could you please rate your overall enjoyment of the session, considering both – the robot and the activity?	*'...The voice is also not pleasant'*
Did you think the amount of interaction was appropriate or should it have been more or less?	*'Passive'* *'Was not interactive'*

Table 5.2: Questionnaire Responses: Awkwardness of Robot

Observation 2

Robot curiosity and attention given to the robot decreases as the session progresses.

It is observed that personalized remarks greatly resonate with the participants. When the robot addresses them by the name during the session, it increases participant excitement, evident through facial expressions, smiles, and exchanged glances among participants. Survey responses also indicate appreciation for personalized comments, suggesting increasing individualization by potentially giving personalized compliments and feedback about their singing itself, as depicted in Table 5.3.

Question	Response
Do you have any other feedback you would like to give to improve the system?	*'Let robot look at the different people if, for example, they sing extra loud or sing incorrectly.'*

Table 5.3: Questionnaire Responses: Preference for Personalized Responses

In addition, personalization during the interaction also combats the decrease in curiosity. It is noticed that participants stop responding to generic compliments and comments that the robot addresses to the entire group. However, when these comments are personalized with their names, it heightens their excitement and happiness, drawing their focus to the robot. With this, we make an inference about the importance of personalization in **Observation 3**.

Observation 3

Personalized comments play an important role in the human-robot connection.

One observation from the video analysis was that participants referred to the robot using human pronouns such as 'he' rather than addressing it as an object with 'it', showing that they addressed the robot as a member of the group, rather than as a mere object. Questionnaire responses echo similar sentiments, as shown with one recommendation being to dress the robot in attire like a shirt, glasses or cap, which will make it more approachable and provide a human-like appearance, as shown in Table 5.4. When the robot asked participants for feedback after the session, another participant expressed the expectation for the robot to sing along with the group, thereby humanizing it and integrating it as a part of the group. These findings are generalized in **Observation 4**.

Question	Response
Reflecting on your experience, could you please rate your overall enjoyment of the session, considering both – the robot and the activity?	*'...If the robot itself starts singing, there might be more commitment...'*
Do you have any other feedback you would like to give to improve the system?	*'Other decoration: e.g.: clothes. he does not communicate. different voice'*

Table 5.4: Questionnaire Responses: Humanization of Robot

Observation 4

Participants personify the robot and express a preference for it to possess greater human-like attributes.

5.2.2. Activity Engagement

Activity engagement, as explained in Subsection 3.2.2 refers to the involvement of the participants in the activity or task itself, which in this case is singing along to music of their choice. Video analysis reveals a remarkably high level of engagement, nearly reaching 100%. This means that all participants remained engaged and sang throughout the entirety of the selected songs. After every song, participants were given the option to continue singing another song or end the session. Both groups opted to sing 4 songs each.

Throughout the session, they displayed emotions of happiness and excitement, which were manifested through visible smiles and enthusiastic engagement, often seen in their spontaneous dancing movements, humming and harmonizing with each other. These expressions serve as visible indicators of their positive experience and enjoyment of the activity. These findings are backed by answers to the subjective questionnaire, examples of which are shown in Table 5.5 where participants indicate a high level of enjoyment and the potential of the activity to improve mood. This video evidence is summarized in **Observation 5**.

Question	Response
Reflecting on your experience, could you please rate your overall enjoyment of the session, considering both – the robot and the activity?	*'Enjoyed'* *'I really enjoyed singing together...'*
Do you believe that engaging in such an activity has the potential to positively influence mood?	*'Singing together always makes me and many others very happy'* *'Pleasant'* *'Can increase mood, but wants to see more interaction from the robot'*

Table 5.5: Questionnaire Responses: Activity Engagement

Observation 5

Participants enjoyed the sing-along activity, evident from their enthusiastic dancing and expressions of excitement.

It's important to consider that participants were recruited from choirs, indicating a pre-existing inclination towards singing. They received a detailed advertisement about the experiment, and only those who were interested chose to sign up, resulting in a selection bias favoring individuals who have a desire to sing. Moreover the experiment was conducted in a lab, and not their home environment. Will they have the same level of engagement otherwise? This would need further investigation and testing. These factors could potentially account for the fixed level of engagement across participants and a lack of variation. This rendered the robot's encouragement unnecessary. Consequently, the robot's efficacy to encourage singing within the group could not be adequately tested.

Observation 6

There is a lack of variation in the engagement among participant, as all participants display a high level of activity engagement.

5.2.3. Social Experience

Social experience and social engagement pertain to how participants interact and behave with each other [18], as defined in Section 1.3. It involves the manner in which they communicate with each other, as well as the attitude and social dynamic they display. This includes aspects like verbal communication, non-verbal cues, gestures and facial expressions.

Upon analysis of these aspects in the video, we can say that the social experience of participants was characterized by active interaction and engagement. During the segment of the session where they must select a song, most participants engaged in eager discussions to make their choices. The power dynamics in both groups were mostly balanced, and no one showed authority over the others. Only one of the participants seemed more passive than the rest while making suggestions, while the others seemed more excited and actively participated in the conversation. We see that though most of the group was actively making suggestions, they ensured that the more quiet participant's choices were also considered.

While singing, one of the groups frequently made eye contact with each other, exchanging smiles, dancing, and swaying along with the music, thus displaying evident expressions of joy and excitement. They would also sometimes talk to each other, either about the music being played or about the actions of the robot. While the other group did not make as much eye contact, they became excited and intrigued when the robot addressed them by their names and offered encouragement, validating their performance. One of the participants also encouraged other members to sing and harmonize along with the music even during sections where there were no lyrics.

These findings and observations lead us to believe that the session was a valuable social experience as they bond and interact as a group. The same is backed by survey responses, as shown in Table 5.6, where all participants admitted that it was a good bonding exercise, and when done over a period of time can sustain connections.

Question	Response
How would you rate the social experience with your group members?	*'Cheerful and social get-together'* *'We didn't know each other, but it was great fun to do this together. there was definitely a positive vibe in the group'*
Do you think the activity could help you in connecting with the others in the group and improve your friendship with them over time?	*'Yes, but then I have to do a follow-up to it'* *'Singing connects'* *'Like many other activities you do with a group, singing together certainly creates bonding'*

Table 5.6: Questionnaire Responses: Social Experience

Observation 7

The robot-moderated session gave the participants a platform and an opportunity to interact with each other and form social bonds.

5.3. Pose Estimation Algorithm: Discussion

As outlined in Section 3.2, an automated approach employing pose estimation is used to find out whether participants respond to the robot's prompts. Unfortunately, during the first group session, the webcam stopped working, and it could not be resolved before the second group came in. Hence, automated pose estimation was carried out only once during the first session. During this, the robot asks users to clap along if they are enjoying the music. As mentioned above, participants do not respond to this prompt. We observe that using the X and Y coordinates of the keypoints and bounding boxes of each detected human, the pose estimation algorithm is correctly able to identify those who were not performing the prompted action. The robot then provides personalized encouragement using the names of the participants (input to the code) for them to join in the activity.

It is essential to consider that the webcam should capture all the participants, with the keypoints of their bodies, especially the wrists and elbows, without which automated pose estimation cannot be performed correctly. In our experiment, the webcam was calibrated before participants came. However, they adjusted their chairs and positions slightly, resulting in the body parts necessary for pose calculation not being visible in the webcam for two participants. This resulted in the pose being calculated for only the other two.

Additionally, before the start of the experiment, the algorithm requires an input of the participant names in the order in which they are sitting. When a disengaged participant is detected, the algorithm matches the position of that individual to the sequence of names input, which it then uses to personalize the encouragement. This is based on the assumption that participants don't change positions during the experiment. In a care home for PwD an older adults, we assume that the physical movement is restricted, and hence they will not move around during the session. However, this needs more testing in a care home to see if the assumption holds.

We choose the robot prompts as probing users to clap their hands and raise them, because it is appropriate in the context of a sing-along. However, for any other group activity these prompts might have to be reconsidered. For example, in an activity that involves painting and art therapy, clapping might not be the most relevant prompt. In this case, other pertinent prompts must be considered, to help identify user engagement in the robot.

Finally, it is important to note that for this experiment we make groups of 4 participants each, a number which can be captured by one webcam. However, in a care home setting, this might not be the case. If there are more participants, the use of multiple webcams might be required to capture all the users for the pose estimation algorithm to work.

Thus, various factors must be considered while using the algorithm to identify engagement. However, since there were not many opportunities to test the algorithm, this technique definitely needs more investigation and testing, especially in a more natural environment than in a lab.

Observation 8

The pose estimation algorithm to detect robot engagement displays good promise, but needs more testing.

5.4. Design Recommendations

Based on the key observations outlined in the previous sections, the following recommendations can be proposed for improving future design and user experience. These recommendations focus on making the system more impactful and effective while making the human-robot interaction more meaningful.

Recommendation 1: Effective Robot Speech: A significant challenge during the measurement of robot engagement was the lack of audibility and comprehensibility of the robot's voice. This issue may have contributed to participants not responding to the robot's prompts and perceiving it as a passive presence in the interaction. Despite efforts to manually lower the volume of the background music when the robot spoke, it remained difficult to understand the robot's dialogues. To address this issue, there are two potential solutions to consider. Firstly, an automated approach could be implemented to adjust the robot and music volume levels dynamically, ensuring that the robot's dialogues are audible amidst the background music. Alternatively, the dialogues could be displayed on the tablet screen of the Pepper robot, providing participants with a visual aid to accompany the auditory instructions.

These recommendations are based on the insights provided in Observations 1 and 2.

Recommendation 2: Organic Interaction: Throughout the interaction, all dialogues were scripted, resulting in a very mechanical interaction. The words of encouragement and prompts too were repetitive, with the robot repeating the same set of lines every now and then. This created an impression of robotic incompetence. Furthermore, there was no flexibility in conversation, making it highly algorithmic. This lack of variation may contribute to the declining attention towards the robot as the session progresses, leading to diminished robot engagement. To address this issue, we propose leveraging Large Language Models (LLMs) to facilitate more natural and authentic conversations. Additionally, incorporating user voice input and formulating responses to it through LLMs could enhance the organic nature of HRI.

These recommendations are provided keeping in mind Observations 1 and 2.

Recommendation 3: Improvements to Pose Estimation: The pose estimation method used to assess user pose and responsiveness to the robot presents a promising step forward. However, several considerations must be acknowledged for its effective implementation. Firstly, it is imperative that all participants remain within the webcam's field of view. While challenging with a group, this can be addressed by positioning the webcam to ensure visibility for everyone or by employing multiple cameras. Failure to achieve this may result in the body keypoints not being visible in the webcam, making pose estimation ineffective. Secondly, the introduction of additional prompts beyond clapping and raising hands, which are detectable by pose estimation, could enhance the system's versatility. Lastly, it is also important to note that the personalized response in pose estimation relies on the condition that all participants maintain consistent seating positions throughout the experiment. This requirement arises from the fact that their positions are pre-defined in the code prior to the experiment. This design was formulated with the understanding that participants in a care home setting may have limited mobility and tend to remain seated in one place, especially considering their age. Consequently, the system addresses each individual by name and offers encouragement based on the position of any participant who may be disengaged. Thus, further testing in a care home is necessary to ascertain whether this design remains effective or if alternative solutions need to be considered. If the positions of the participants is noticed to vary during the session, face recognition could be employed in conjunction with pose estimation to provide added flexibility.

These recommendations work upon the insights obtained in Observation 8.

Recommendation 4: Personalize the interaction:

Personalization plays an important role, particularly in group HRI. In this experiment, personalized feedback was integrated to address disengagement in participants, wherein the robot would address individuals by name and provide encouragement to re-engage. Even during instances of enthusiastic participation, the robot offered personalized motivational support, a gesture that was well-received by users, as evidenced by both video analysis and survey responses.

Apart from this, incorporating face recognition to greet participants with eye contact can enhance user experience. Further, including a memory model to remember participant behaviour and past conversations can help the robot initiate interactions, making it more meaningful and helping create a bond with the user.

This could help even when the initial excitement and curiosity of interacting with a robot wears off, and can support sustaining the human-robot relationship over a longer period of time. Entering the realm of multimedia recommendation, even remembering the user's choice of music and accordingly suggesting and playing songs can help tailor the playlist to the user's preferences.

These suggestions could help improve on the findings in Observations 2, 3 and 4.

Recommendation 5: Humanize the robot's demeanor:
While this thesis attempted to incorporate the human characteristic of engagement detection, this can be further enhanced. Dressing the robot in clothes and the use of eye contact can make it more approachable and likeable. Participants found the voice mechanical and unpleasant, suggesting that a transition to a more human-like voice could be well-received. As mentioned earlier, the ability of conversing with the users naturally can also be a major advantage in the context of acceptance into the human-robot teams. It is evident that participants already refer to the robot with human pronouns, and suggested improvements might facilitate the robot's deeper integration into the group.

Although the feedback obtained suggests the need to humanize the robot, it could lead to unrealistic expectations. For example, even during this research, the participants expected the robot to sing, which may be beyond its current capabilities, resulting in disappointment or reduced trust towards the robot. There is also research which points out the negative aspects of humanization of robots [8]. It shows that people accept robots as machines for task-based roles, but are uncomfortable with the idea of a humanoid robot which performs social interactions. Such robots might evoke feelings of discomfort and eeriness and could be perceived as a threat to human identity. These considerations should be taken into account while designing a humanoid robot. Care must be taken to strike a balance between human-like and non-human characteristics to help manage user expectations.

These suggestions work upon the feedback summarized in Observation 4.

6
Discussion and Conclusion

6.1. Discussion

Analysis of the session videos gives us several insights into preferences of older adults in terms of HRI and recommendations for future design. Firstly, the questionnaire results show that the robot is found to be incompetent, passive and unresponsive, with participants expecting more interaction from it. The goal of the thesis was to build a robot that can guide and motivate users through a sing-along session. However, participants expect the robot to be more involved in the activity. Participants also express the desire of the robot to sing with them, signifying that they want the robot to be an active participant, rather than a simple bystander or an activity moderator. Thus, a robot that can integrate into the group, interact organically with others and possibly even sing, would likely receive more favourable responses. These changes will be consequential in building trust and establishing long-lasting relations with users.

Secondly, we see that curiosity plays an important role in the attention given to the robot, as also noted in the questionnaire responses. However, we see that this decreases as they get accustomed to the robot's presence. Attention to the robot decreases as time passes, and their intrigue is only maintained by personalized comments and remarks. Hence, personalization is an effective method to keep the users engaged in the robot as well as the activity.

While participants note that their initial intrigue in the robot encouraged them to sing, they did not feel that the robot provided positive encouragement with verbal comments or enhanced the overall activity. It is essential to acknowledge that the participants already had an inclination toward singing, indicating a selection bias. Due to this, the efficacy of the robot in motivating users could not be gauged. Can the robot re-engage users? This question remains unanswered. Further testing carried out over a longer period of time is required to evaluate the its effectiveness in sustaining user engagement.

Finally, we see that the activity can help improve social bonds with peers. Video observations indicate that while singing, participants made constant eye contact with each other and motivated each other to sing and dance. The decision-making process about song choices was also seen to be inclusive. Everyone contributed to the discussion, making it a collective choice. Preferences of the more passive members were taken into consideration too, showing a sense of group collaboration. Survey responses also indicate that the activity could improve social engagement among members when done over a long period of time. Building on these findings, we can say that it has potential in improving social relationships between peers, providing an opportunity for them to interact with each other, thereby improving QoL when deployed in a care home.

6.2. Conclusion

In this thesis, we designed a robot that can interact with groups of people and guide them through a sing-along activity, while providing personalized motivation and encouragement to improve their mood and help them connect with their peers. The system is devised for individuals with dementia who reside in a care home, and aims to improve their social engagement and quality of life by providing a platform and an opportunity for them to interact with fellow residents. Additionally, the employment of the robot in care homes to guide and assist sing-along activities seeks to reduce the workload on caregivers who would otherwise conduct the activity. This allows them more time to attend to the personal needs of the residents and manage their day-to-day activities. The robot performs engagement detection in real-time, and attempts to inspire those who are not participating in the activity with personalized encouragement. This pushes participants to join in the activity, potentially boosting their morale, improving their mood and giving them a chance to interact with others.

A new engagement detection technique is proposed, which makes use of the presence of the robot to understand whether the user is involved and engaged. This represents a shift from the traditional passive paradigms of detecting engagement, wherein the robot's role is not leveraged. In the proposed methodology, the robot interactively probes the user to perform certain actions while they are singing. These actions could be either clapping along to the music or raising their hands. We then detect, using pose estimation, whether these prompts are being carried out by the users. This method provides insights into the responsiveness of users to the robot, and potentially their involvement in the activity. This approach shows promise for use in future research, but needs more testing to make it robust.

Reflecting on the first research question (RQ1), which delves into the acceptance of the robot in older adult communities and their perceptions about it, we note that there are mixed perceptions about the robot. Majority of the participants view the robot as incompetent, and expect more conversation and interaction with it, preferably personalized. Some also find it unpleasant to interact with due to its mechanical voice. Additionally, it may be deemed strange, with a passive presence, as some participants do not notice it much. In contrast, some users find the robot intriguing and mention that its inclusion in the activity makes it enjoyable, as the robot guides them through the process. They appreciate the positive comments made by the robot and like it when the robot personalizes the interaction.

With respect to the second research question (RQ2) regarding enhancement of social engagement in a robot-guided group sing-along activity, findings from the subjective questionnaire and videos suggest that such an activity with a robot enhances the social experience and helps them connect with the group. Not only does the robot conduct the session, but it also provides an opportunity for them to communicate and engage with with each other. This shows that the robot-moderated activity has potential to improve human-human interactions within groups, offering promising implications, particularly in care homes catering to individuals with dementia.

While this research yields encouraging results, it definitely has scope for improvement. Designing for a specific user group requires exploration into the dynamics and behaviour of the group, with numerous feedback cycles to refine the system. The study provides an initial round of feedback, laying the groundwork and paving the way for future enhancements.

6.3. Limitations

Upon reflection of the methodology and experiments, several limitations and areas of improvement are noticed in the experiments and the implemented prototype. Due to time constraints, only one session per group could be conducted. Ideally, the intervention should span over an extended period, perhaps for a few months, with multiple sessions during this duration. This enables us to accurately gauge alterations in mood and social engagement, rather than relying on data from a single session. Furthermore, the number of groups was quite small. This was the result of restricted access to older adults and logistical challenges, as they were required to travel to TU Delft for participation, which may have deterred some individuals from attending. The small sample size leads to unreliable results and limited generalizability.

Secondly, participant recruitment was carried out by sending posters of the experiment to communities of older adults, allowing them to decide whether to participate. This led to a significant selection bias among participants. Since they were aware the experiment involved singing, only highly motivated individuals with an interest in singing attended. Therefore, there was a lack of variation in activity engagement as everyone sang nearly 100% of all songs, and prevented testing of the condition where the robot would encourage participants. Thus, insights into the efficacy of the robot's encouragement could not be obtained, as a natural scenario from a care home was not emulated.

Further, older adults might not be the most accurate substitution of the target audience, which is PwD. This is because they do not share the same behaviour and decision making skills. Though both groups might face similar age-related challenges, there might be a distinct difference in the way they interact with and respond to their environment. This difference in behaviour is attributed to the cognitive decline associated with dementia. The session and prototype is designed for those in a care home, who have limited interaction with fellow residents, a circumstance not entirely replicated by older adults. This can impact the reliability and effectiveness of the results and feedback obtained, making it important to test the prototype with the intended audience.

The conversations and robot dialogues were scripted, with little to no scope for deviation from the predetermined script. Some of the robot responses were given by the human operator, limiting the autonomous behavior of the robot. This rigid approach made the robot seem inflexible, with participants not paying much attention to it. The lack of flexibility likely contributed to the perceived incompetence. Moreover, since the experiment was in Dutch, all interactions within the group occurred in Dutch. However, since the human operator (I) did not speak the language, they were unable to fully grasp the conversations within the group, limiting the ability to prompt an appropriate response from the robot.

Additionally, during the experiment, there were certain unforeseen circumstances. The session used 4 microphones, however they interfered with each other preventing them from operating simultaneously. This also caused a high pitched noise from the speaker to which the microphone was connected, causing frustration in the participants. The webcam too stopped working during the experiment, hindering automated pose estimation. This compelled us to fall back on manual intervention for the same.

Finally, the robot platforms posed issues, with the robot behaving unexpectedly at times. The SIC framework was still under development, with limited documentation for the existing features. The attempt to perform robot gestures and body poses with SIC led to unpredictable behaviour from the robot, like turning it's head to one side all of a sudden. This led to reduced efficiency, as the entire code had to be explored to understand how to use any function.

6.4. Future Work

With the field of HRI with individuals with dementia under constant research and development, there are various avenues to enhance the system, merge it with other ongoing research and make changes based on the results and consequent observations.

Firstly, the conversation responses should be automated with LLMs, making the interaction more natural and less mechanical. As noted in the feedback obtained during the experiments, the robot is perceived as incapable and unresponsive. To counter this and make the robot more interactive, it should be able to engage in conversations organically. To achieve this, the robot needs the capability to transcribe the user's speech to text, requiring the use of more robust microphones to accurately capture the voice. The transcribed text must be input to the LLM, and the robot then responds with the output generated.

In this thesis, activity engagement detection was facilitated by a WoZ, where a human operator detected whether a participant was singing or not and accordingly prompted the robot response. To streamline this process and possibly eliminate human intervention, this must be automated. This can be done by using unidirectional microphones for the users to sing into and then monitoring signals obtained to identify which participants are singing. If some users are shy and choose to sing without a microphone, other techniques would have to be researched for the same.

One major drawback of this study was the inability to perform evaluations with people with dementia in a care home. For more accurate results and observations, the prototype must be tested with them. Such testing could offer a different and new perspective on the robot's role and effectiveness. Further, these sessions should be conducted over an extended duration to assess their long term impact. A single session is insufficient to accurately judge the usefulness of the session in bonding the group.

It must also be examined whether the introduction of such a robot can assist caregivers in nursing homes. While testing the robot in a care home gives us insights into how it helps persons with dementia, we must also think from the perspective of formal caregivers. Conducting interviews and focus groups with them can provide an alternative outlook, and give design recommendations that we may not have previously considered. Additionally, this will aid in evaluating the practical usefulness of such an intervention and its effectiveness in supporting them and their activities.

The robot guides the user through the process of selecting a song, and the group must interact with each other to select it. Incorporating mediating abilities in case there is a conflict within the group can facilitate resolution and humanize the robot. It could also support decision making processes by assisting users in selecting a song that best suits the preferences and mood of the group.

There is several ongoing research on the relationship between music and memory, looking into how music can stimulate memory and recall certain emotions associated with this music [33][34]. In the future, we can integrate these findings into the work of the current thesis and its methodology. By curating specialized playlists for the group and each of its participants, we seek to evoke emotional responses from them and personalize the experience.

Such a robot can also be enhanced to play music-related games, not only incorporating the benefits of music therapy, but also challenging them to improve upon decision making, strategizing and problem solving skills. Interactive games among the group can stimulate cognitive function and encourage active participation. Thus, combining music and games can provide a multifaceted approach to building an SAR to be deployed in a care home.

Thus, there are numerous paths to explore and advance research in the field of SARs for PwD forward. While this thesis provides a first round of feedback for development of a robot-guided sing-along, it still has a long way to go. Iterative development along with continued research and testing will be key to realizing the full potential of robot-moderated activities in enhancing QoL of PwD.

A

Human-Robot Interaction Dialogue

Step 1: Introduction

Speaker	Dialog	Additional Information
Robot	*(Waves at participants)* "Hi Jack and Jane! I'm Pepper! I trust you are having a good day. Would you like to participate in a game? I will explain the rules of the game. I have many songs to choose from, and I will help you make the choice too. Once you all have chosen a song, let me know and I'll play it for you. You then sing along to your favorite music and have fun! When the song is over, please let me know if you would like to sing another song. Ready? Let's begin!" *(Enthusiastic gesture)*	The names Jack and Jane are placeholder names, and are replaced with actual participant names during the experiment

Table A.1: Introduction: Dialog Flow

Step 2: Music Choices

Speaker	Dialog	Additional Information
Robot	"Do you want to sing a Dutch or English song? These choices are also displayed on the screen."	
Participants	*(discuss among themselves)* "We would like to sing an English song!"	Actual participant conversations may differ.
Robot	"Choose the genre you want to sing. The choices are displayed on the screen."	
Participants	*(discuss among themselves)* "We would like to sing a Pop song!"	
Robot	"I like your choice in music! Choose the song you want to sing from the list shown on the screen. Discuss it with your friends and let me know which song you want to sing along to."	A Spotify playlist is opened on the screen.
Participants	*(discuss among themselves)* "Can you play Summer of '69?"	

Table A.2: Music Choices: Dialog Flow

Step 3: Singing

	Speaker	Dialog	Additional Information
	Robot	*(Music starts playing)* “Come sing along!”	
Robot Engagement	Robot	“If you’re having a good time, clap along to the rhythm.”	WoZ can choose one of these prompts during the song
	Robot	“Raise your hands and groove to the music” *(Raises hands)*	
	Participant 1 (Jack)	*(Does not respond to robot prompt)*	
	Robot	“Hey Jack, come join in with the group!”	Personalized motivation for participants not performing the prompt.
Activity Engagement	Robot	“Wow, you are singing so well.”	Can be addressed to the group or one person depending on who is singing (Any of these dialogues can be chosen by the WoZ)
	Robot	“You all are amazing singers.”	
	Robot	“Your voices complement each other perfectly. It’s such a pleasure to listen to.”	
	Robot	“You sing so well that I feel like dancing.”	
	Robot	“Don’t be shy, sing with heart and soul.”	Can be personalized to motivate someone who is not singing (Any of these dialogues can be chosen by the WoZ)
	Robot	“Come on join in.”	
	Robot	“We would love to listen to your voice, sing loud.”	
	Robot	“Join in with everyone.”	

Table A.3: Singing: Dialog Flow

Step 4: Next Song

Speaker	Dialog	Additional Information
Robot	*(Song stops playing)* “Thanks for singing along to this song! If you’re having fun, would you like to play another song?”	
Participants	*(discuss among themselves)* “Let’s sing another song!”	
Robot	“I’m glad to know you’re enjoying it!”	If participants choose to sing another song, flow moves to Step 2.
Participants	*(discuss among themselves)* ‘I think we would like to stop for now”	If participants stop, flow moves to Step 5.

Table A.4: Next Song: Dialog Flow

Step 5: Closing of the Session

Speaker	Dialog	Additional Information
Robot	“Thank you, everyone, for participating. I enjoyed hearing you sing, you are certainly a fantastic and enthusiastic bunch of people.”	Depending on the enthusiasm of the group, the WoZ chooses one of these dialogues for the robot to say.
Robot	“Thank you, everyone, for participating. I enjoyed hearing you sing. I hope you won’t be so shy next time we meet!”	
Robot	“Did you have fun? How did you feel about singing along with me?”	
Participants	“I definitely had fun singing!”	
Robot	“Thank you for joining me today. I hope to see you soon.”	

Table A.5: Closing of the Session: Dialog Flow

B

Rubric for qualitative video analysis

To analyze video data, the following protocols must be answered, and then analyzed for every participant of the session.

Phase	Question	Possible Answers
Introduction	What is the participant looking at?	Robot, Screen, Other Participants, Facilitator, Down, other
	Is the participant interacting with other participants?	Yes, No. If yes, is it verbally or behaviourally?
	Mood of the Participant	Pleasure, Frustration, Fear, Sadness, Boredom, Neutral
	Other Feedback	Observations not mentioned above

Table B.1: Phase 1: Protocol to analyze video data

Phase	Question	Possible Answers
Song Selection	What is the participant looking at?	Robot, Screen, Other Participants, Facilitator, Down, other
	Is the participant interacting with other participants?	Yes, No. If yes, is it verbally or behaviourally?
	Are they actively or passively participating in song choices?	Actively, Passively
	Mood of the Participant	Pleasure, Frustration, Fear, Sadness, Boredom, Neutral
	Other Feedback	Observations not mentioned above

Table B.2: Phase 2: Protocol to analyze video data

Phase	Question	Possible Answers
Singing	What is the participant looking at?	Robot, Screen, Other Participants, Facilitator, Down, other
	Is the participant interacting with other participants? Does this include encouraging them to sing?	Yes, No. If yes, is it verbally or behaviourally?
	How much of the song do they sing?	Did not sing at all, Sang very little, Sang half the song, Sang most of it, Sang all of it
	Do they respond to robot prompts?	Yes, No. If no, why not?
	Did the prompts given by the robot encourage them to start singing? If yes, what was the percentage of times it happened?	Yes, No
	Mood of the Participant	Pleasure, Frustration, Fear, Sadness, Boredom, Neutral
	Other Feedback	Observations not mentioned above

Table B.3: Phase 3: Protocol to analyze video data

Phase	Question	Possible Answers
Final	What is the participant looking at?	Robot, Screen, Other Participants, Facilitator, Down, other
	Is the participant interacting with other participants? If yes, is it verbally or behaviourally?	Yes, No
	Do they respond to robot prompts when asked about the experience?	Yes, No
	What feedback did they give to the robot?	Answer this depending on the answer given in the video
	Mood of the Participant	Pleasure, Frustration, Fear, Sadness, Boredom, Neutral
	Other Feedback	Observations not mentioned above

Table B.4: Phase 4: Protocol to analyze video data

C

Questionnaire 1: RoSAS

The following questionnaire is a validated questionnaire to measure social perception of robots. There are 18 characteristics, for which the robot can be rated on a scale of 1-9. You must answer whether you think these characteristics are associated with the robot you interacted with.

Using the scale provided, how closely are the words below associated with the robot you just interacted with?

	Definitely not associated								Definitely associated
Happy	1	2	3	4	5	6	7	8	9
Feeling	1	2	3	4	5	6	7	8	9
Social	1	2	3	4	5	6	7	8	9
Organic	1	2	3	4	5	6	7	8	9
Compassionate	1	2	3	4	5	6	7	8	9
Emotional	1	2	3	4	5	6	7	8	9
Capable	1	2	3	4	5	6	7	8	9
Responsive	1	2	3	4	5	6	7	8	9
Interactive	1	2	3	4	5	6	7	8	9
Reliable	1	2	3	4	5	6	7	8	9
Competent	1	2	3	4	5	6	7	8	9
Knowledgeable	1	2	3	4	5	6	7	8	9
Scary	1	2	3	4	5	6	7	8	9
Strange	1	2	3	4	5	6	7	8	9
Awkward	1	2	3	4	5	6	7	8	9
Dangerous	1	2	3	4	5	6	7	8	9
Awful	1	2	3	4	5	6	7	8	9
Aggressive	1	2	3	4	5	6	7	8	9

D

Questionnaire 2: Subjective Questionnaire

The following questions have been developed to gain initial feedback on the robot and experiment. It includes questions about the overall experience, the social experience, the role and interaction with the robot and some general feedback. We are interested in understanding your perspective, so please feel free to provide specific details or examples to support your answer.

Overall Experience

Reflecting on your experience, could you please rate your overall enjoyment of the session, considering both – the robot and the activity?

Did not enjoy at all				Enjoyed to the fullest
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Do you believe that engaging in such an activity has the potential to positively influence mood?

Cannot improve mood at all				Can improve mood a lot
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Social Experience

How would you rate the social experience with your group members?

Did not like performing the activity in a group				Absolutely loved performing the activity in a group
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Do you think the activity could help you in connecting with the others in the group and improve your friendship with them over time?

The activity does not help at all in connecting with other group members				The activity surely helps in connecting with other group members
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Robot Role and Interaction

How would you rate the role played by the robot (actions and behaviour) throughout the experience?

Disliked the robot				Liked the robot
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Did the robot actively encourage participation?

Did not encourage participation at all				Encouraged participation a lot
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Did you think the amount of interaction was appropriate or should it have been more or less?

Robot should have interacted lesser				Robot should have interacted more
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Did the robot enhance the activity for you?

Did not enhance my experience at all				Enhanced my experience quite a lot
1	2	3	4	5

Can you provide a detailed explanation for why you gave the above rating?

Overall Feedback

Do you have any other feedback you would like to give to improve the system?

Bibliography

[1] G. Terzopoulos and M. Satratzemi, "Voice assistants and smart speakers in everyday life and in education," *Informatics in Education*, vol. 19, pp. 473–490, Sep. 2020. DOI: 10.15388/infedu.2020.21.

[2] C. Gyles, "Robots in medicine," *Can Vet J*, vol. 60, no. 8, pp. 819–820, Aug. 2019.

[3] E. Konijn, M. Smakman, and R. van den Berghe, "Use of robots in education," pp. 1–8, Sep. 2020. DOI: 10.1002/9781119011071.iemp0318.

[4] M. Ghafurian, C. Ellard, and K. Dautenhahn, "Social companion robots to reduce isolation: A perception change due to covid-19," in *Human-Computer Interaction – INTERACT 2021*, C. Ardito, R. Lanzilotti, A. Malizia, *et al.*, Eds., Cham: Springer International Publishing, 2021, pp. 43–63, ISBN: 978-3-030-85616-8.

[5] A. Tapus, M. J. Mataric, and B. Scassellati, "Socially assistive robotics [grand challenges of robotics]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 35–42, 2007. DOI: 10.1109/MRA.2007.339605.

[6] D. Feil-Seifer and M. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, 2005, pp. 465–468. DOI: 10.1109/ICORR.2005.1501143.

[7] C. Yu, A. Sommerlad, L. Sakure, and G. Livingston, "Socially assistive robots for people with dementia: Systematic review and meta-analysis of feasibility, acceptability and the effect on cognition, neuropsychiatric symptoms and quality of life," *Ageing Research Reviews*, vol. 78, p. 101 633, 2022, ISSN: 1568-1637. DOI: https://doi.org/10.1016/j.arr.2022.101633.

[8] J.-C. Giger, N. Piçarra, P. Alves-Oliveira, R. Oliveira, and P. Arriaga, "Humanization of robots: Is it really such a good idea?" *Human Behavior and Emerging Technologies*, vol. 1, no. 2, pp. 111–123, 2019. DOI: https://doi.org/10.1002/hbe2.147.

[9] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 143–166, 2003, Socially Interactive Robots, ISSN: 0921-8890. DOI: https://doi.org/10.1016/S0921-8890(02)00372-X.

[10] D. Silvera-Tawil and C. R. Yates, "Socially-assistive robots to enhance learning for secondary students with intellectual disabilities and autism," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 838–843. DOI: 10.1109/ROMAN.2018.8525743.

[11] D. Macis, S. Perilli, and C. Gena, "Employing socially assistive robots in elderly care," in *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, ser. UMAP '22 Adjunct, Barcelona, Spain: Association for Computing Machinery, 2022, pp. 130–138, ISBN: 9781450392327. DOI: 10.1145/3511047.3537687.

[12] *What is dementia? symptoms, causes and treatment*, https://www.alz.org/alzheimers-dementia/what-is-dementia.

[13] W. Moyle, U. Kellett, A. Ballantyne, and N. Gracia, "Dementia and loneliness: An australian perspective," *Journal of clinical nursing*, vol. 20, pp. 1445–53, Mar. 2011. DOI: 10.1111/j.1365-2702.2010.03549.x.

[14] L. Adlbrecht, S. Bartholomeyczik, C. Hildebrandt, and H. Mayer, "Social interactions of persons with dementia living in special care units in long-term care: A mixed-methods systematic review," *Dementia*, vol. 20, p. 147 130 122 091 993, Apr. 2020. DOI: 10.1177/1471301220919937.

[15] C. Brown Wilson, "Developing community in care homes through a relationship-centred approach," *Health & social care in the community*, vol. 17, pp. 177–86, Apr. 2009. DOI: 10.1111/j.1365-2524.2008.00815.x.

[16] *Dementia: World health organization*, https://www.who.int/news-room/fact-sheets/detail/dementia.

[17] L. Moshkina, S. Trickett, and J. G. Trafton, “Social engagement in public places: A tale of one robot,” in *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2014, pp. 382–389.

[18] W. Avison, J. McLeod, and B. Pescosolido, *Mental Health, Social Mirror*. Jan. 2007, pp. 1–471, ISBN: 978-0-387-36319-6. DOI: 10.1007/978-0-387-36320-2.

[19] M. Luo, D. Ding, A. Bauman, J. Negin, and P. Phongsavan, “Social engagement pattern, health behaviors and subjective well-being of older adults: An international perspective using who-sage survey data,” *BMC Public Health*, vol. 20, Jan. 2020. DOI: 10.1186/s12889-019-7841-7.

[20] K. R. Krueger, R. S. Wilson, J. M. Kamenetsky, L. L. Barnes, J. L. Bienias, and D. A. Bennett, “Social engagement and cognitive function in old age,” *Exp. Aging Res.*, vol. 35, no. 1, pp. 45–60, Jan. 2009.

[21] B. D. James, R. S. Wilson, L. L. Barnes, and D. A. Bennett, “Late-life social activity and cognitive decline in old age,” *J. Int. Neuropsychol. Soc.*, vol. 17, no. 6, pp. 998–1005, Nov. 2011.

[22] K. D. Moore and R. Verhoef, “Special care units as places for social interaction: Evaluating an scu’s social affordance,” *American Journal of Alzheimer’s Disease*, vol. 14, no. 4, pp. 217–229, 1999. DOI: 10.1177/153331759901400406.

[23] B. Hanratty, D. Stow, D. Collingridge Moore, N. K. Valtorta, and F. Matthews, “Loneliness as a risk factor for care home admission in the English Longitudinal Study of Ageing,” *Age and Ageing*, vol. 47, no. 6, pp. 896–900, Jul. 2018, ISSN: 0002-0729. DOI: 10.1093/ageing/afy095.

[24] J. Cohen-Mansfield, “The impact of group activities and their content on persons with dementia attending them,” *Alzheimer’s research & therapy*, vol. 10, p. 37, Apr. 2018. DOI: 10.1186/s13195-018-0357-z.

[25] K. Cherry, E. Walker, J. Brown, *et al.*, “Social engagement and health in younger, older, and oldest-old adults in the louisiana healthy aging study (lhas),” *Journal of applied gerontology : the official journal of the Southern Gerontological Society*, vol. 32, pp. 51–75, Feb. 2013. DOI: 10.1177/0733464811409034.

[26] H. S. Aasgaard, B. T. Landmark, and S. Öresland, “Healthcare personnel’s experiences of enhancing social community for people living in dementia special care units,” *Nordic Journal of Nursing Research*, vol. 37, no. 4, pp. 186–193, 2017. DOI: 10.1177/2057158517693534.

[27] M. Rahimi, R. Fadayevatan, and H. A. Abedi, “Care instability in nursing homes; a qualitative study,” *Iran. Red Crescent Med. J.*, vol. 18, no. 2, Feb. 2016.

[28] N. Kandelman, T. Mazars, and A. Levy, “Risk factors for burnout among caregivers working in nursing homes,” *Journal of clinical nursing*, vol. 27, May 2017. DOI: 10.1111/jocn.13891.

[29] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung, “Robots in groups and teams: A literature review,” *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW2, Oct. 2020. DOI: 10.1145/3415247.

[30] A. Wisowaty, “Group human-robot interaction: A review,” 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:210941628.

[31] C. Oertel, G. Castellano, M. Chetouani, *et al.*, “Engagement in human-agent interaction: An overview,” *Frontiers in Robotics and AI*, vol. 7, 2020, ISSN: 2296-9144. DOI: 10.3389/frobt.2020.00092.

[32] M. Bleibel, A. Cheikh, N. Said Sadier, and L. Abbas, “The effect of music therapy on cognitive functions in patients with alzheimer’s disease: A systematic review of randomized controlled trials,” *Alzheimer’s Research & Therapy*, vol. 15, pp. 1–10, Mar. 2023. DOI: 10.1186/s13195-023-01214-9.

[33] A. Raglio and O. Oasi, “Music and health: What interventions for what results?” *Frontiers in Psychology*, vol. 6, 2015, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.00230.

[34] H. Platel, J.-C. Baron, B. Desgranges, F. Bernard, and F. Eustache, “Semantic and episodic memory of music are subserved by distinct neural networks,” *Neuroimage*, vol. 20, no. 1, pp. 244–256, Sep. 2003.

[35] S. E. Osman, V. Tischler, and J. Schneider, “Singing for the brain: A qualitative study exploring the health and well-being benefits of singing for people with dementia and their carers,” *Dementia*, vol. 15, no. 6, pp. 1326–1339, Nov. 2016. DOI: 10.1177/1471301214556291.

[36] A. Miyazaki and H. Mori, “Frequent karaoke training improves frontal executive cognitive skills, tongue pressure, and respiratory function in elderly people: Pilot study from a randomized controlled trial,” *International Journal of Environmental Research and Public Health*, vol. 17, p. 1459, Feb. 2020. DOI: 10.3390/ijerph17041459.

[37] B. Dubois, A. Slachevsky, I. Litvan, and B. Pillon, "The FAB," *Neurology*, vol. 55, no. 11, pp. 1621–1626, 2000. DOI: 10.1212/WNL.55.11.1621.

[38] M. Satoh, T. Yuba, K.-i. Tabei, *et al.*, "Music Therapy Using Singing Training Improves Psychomotor Speed in Patients with Alzheimer's Disease: A Neuropsychological and fMRI Study," *Dementia and Geriatric Cognitive Disorders Extra*, vol. 5, no. 3, pp. 296–308, Sep. 2015, ISSN: 1664-5464. DOI: 10.1159/000436960.

[39] E. Mitchell, "Community music therapy and participatory performance: Case study of a coffee house," *Voices: A World Forum for Music Therapy*, vol. 19, no. 1, Jan. 2019. DOI: 10.15845/voices.v19i1.2701. [Online]. Available: https://voices.no/index.php/voices/article/view/2701.

[40] J. Miller and T. McDaniel, "I enjoyed the chance to meet you and i will always remember you: Healthy older adults' conversations with misty the robot," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 914–918. DOI: 10.1109/HRI53351.2022.9889523.

[41] S. T. Hansen, H. J. Andersen, and T. Bak, "Practical evaluation of robots for elderly in Denmark — an overview," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010, pp. 149–150. DOI: 10.1109/HRI.2010.5453220.

[42] F. Carros, J. Meurer, D. Löffler, *et al.*, "Exploring human-robot interaction with the elderly: Results from a ten-week case study in a care home," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12, ISBN: 9781450367080. DOI: 10.1145/3313831.3376402.

[43] Y.-C. Lin, J. Fan, J. A. Tate, N. Sarkar, and L. C. Mion, "Use of robots to encourage social engagement between older adults," *Geriatr Nurs*, vol. 43, pp. 97–103, Nov. 2021.

[44] A. Leung, I. Zhao, S. Lin, and T. Lau, "Exploring the presence of humanoid social robots at home and capturing human-robot interactions with older adults: Experiences from four case studies," *Healthcare*, vol. 11, p. 39, Dec. 2022. DOI: 10.3390/healthcare11010039.

[45] F. Martín, C. Durán, J. Plaza, *et al.*, "Robots in therapy for dementia patients," *Journal of Physical Agents (JoPha)*, vol. 7, pp. 49–56, Jan. 2013. DOI: 10.14198/JoPha.2013.7.1.07.

[46] J. Broekens, M. Heerink, and H. Rosendal, "Assistive social robots in elderly care: A review," *Gerontechnology*, vol. 8, pp. 94–103, Apr. 2009. DOI: 10.4017/gt.2009.08.02.002.00.

[47] S. M. Villaroya, J. J. Gamboa-Montero, A. Bernardino, M. Maroto-Gómez, J. C. Castillo, and M. Á. Salichs, "Real-time engagement detection from facial features," in *2022 IEEE International Conference on Development and Learning (ICDL)*, 2022, pp. 231–237. DOI: 10.1109/ICDL53763.2022.9962228.

[48] A. Gupta, R. Jaiswal, S. Adhikari, and V. Balasubramanian, "DAISEE: dataset for affective states in e-learning environments," *CoRR*, vol. abs/1609.01885, 2016. arXiv: 1609.01885.

[49] H. Abdollahi, M. Mahoor, R. Zandie, J. Sewierski, and S. Qualls, "Artificial emotional intelligence in socially assistive robots for older adults: A pilot study," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1–1, Jan. 2022. DOI: 10.1109/TAFFC.2022.3143803.

[50] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, Dec. 1980. DOI: 10.1037/h0077714.

[51] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP natural Language Processing Toolkit," Jan. 2014. DOI: 10.3115/v1/P14-5010.

[52] I. A. Valagkouti, C. Troussas, A. Krouska, M. Feidakis, and C. Sgouropoulou, "Emotion recognition in human-robot interaction using the NAO robot," *Computers*, vol. 11, no. 5, 2022, ISSN: 2073-431X. DOI: 10.3390/computers11050072.

[53] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708. DOI: 10.1109/CVPR.2014.220.

[54] R. Oliveira, P. Arriaga, and A. Paiva, "Human-robot interaction in groups: Methodological and research practices," *Multimodal Technologies and Interaction*, vol. 5, no. 10, 2021, ISSN: 2414-4088. DOI: 10.3390/mti5100059.

[55] M. Quiroz, R. Patiño, J. Diaz-Amado, and Y. Cardinale, "Group emotion detection based on social robot perception," *Sensors*, vol. 22, p. 3749, May 2022. DOI: 10.3390/s22103749.

[56] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," vol. 1, Feb. 2001, pp. I–511, ISBN: 0-7695-1272-0. DOI: 10.1109/CVPR.2001.990517.

[57] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2015. arXiv: 1409.1556 [cs.CV].

[58] H. Gunes and N. Churamani, "Affective computing for human-robot interaction research: Four critical lessons for the hitchhiker," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 1565–1572. DOI: 10.1109/RO-MAN57019.2023.10309450.

[59] A. Ben-Youssef, G. Varni, S. Essid, and C. Clavel, "On-the-fly detection of user engagement decrease in spontaneous human–Robot interaction using recurrent and deep neural networks," *International Journal of Social Robotics*, vol. 11, no. 5, pp. 815–828, Dec. 2019, ISSN: 1875-4805. DOI: 10.1007/s12369-019-00591-2.

[60] A. Ben-Youssef, C. Clavel, S. Essid, M. Bilac, M. Chamoux, and A. Lim, "UE-HRI: A new dataset for the study of user engagement in spontaneous human-robot interactions," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI '17, Glasgow, UK: Association for Computing Machinery, 2017, pp. 464–472, ISBN: 9781450355438. DOI: 10.1145/3136755.3136814.

[61] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg, "From real-time attention assessment to "with-me-ness" in human-robot interaction," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 157–164. DOI: 10.1109/HRI.2016.7451747.

[62] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874. DOI: 10.1109/CVPR.2014.241.

[63] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009, ISSN: 1532-4435.

[64] S. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the engagement with social robots," *International Journal of Social Robotics*, vol. 7, Aug. 2015. DOI: 10.1007/s12369-015-0298-7.

[65] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models," in *BMVC 2006 - Proceedings of the British Machine Vision Conference 2006|BMVC - Proc. Br. Mach. Vis. Conf.*, United Kingdom: BMVA Press, 2006, pp. 929–938.

[66] F. Del Duchetto, P. Baxter, and M. Hanheide, "Are you still with me? Continuous engagement assessment from a robot's point of view," *Frontiers in Robotics and AI*, vol. 7, 2020, ISSN: 2296-9144. DOI: 10.3389/frobt.2020.00116.

[67] P. Alves-Oliveira, P. Sequeira, F. S. Melo, G. Castellano, and A. Paiva, "Empathic robot for group learning: A field study," *CoRR*, vol. abs/1902.01800, 2019. arXiv: 1902.01800.

[68] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Trans. Inf. Syst.*, vol. 2, no. 1, pp. 26–41, Jan. 1984, ISSN: 1046-8188. DOI: 10.1145/357417.357420.

[69] L. Bechade, G. Dubuisson-Duplessis, G. Pittaro, M. Garcia, and L. Devillers, "Towards metrics of evaluation of pepper robot as a social companion for the elderly: 8th international workshop on spoken dialog systems," in Jan. 2019, pp. 89–101, ISBN: 978-3-319-92107-5. DOI: 10.1007/978-3-319-92108-2_11.

[70] F. Rietz, A. Sutherland, S. Bensch, S. Wermter, and T. Hellström, "Woz4u: An open-source wizard-of-oz interface for easy, efficient and robust hri experiments," *Frontiers in Robotics and AI*, vol. 8, p. 668 057, Jul. 2021. DOI: 10.3389/frobt.2021.668057.

[71] L. Riek, "Wizard of Oz studies in HRI: A systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, Aug. 2012. DOI: 10.5898/JHRI.1.1.Riek.

[72] N. M. Fraser and G. Gilbert, "Simulating speech systems," *Computer Speech & Language*, vol. 5, no. 1, pp. 81–99, 1991, ISSN: 0885-2308. DOI: https://doi.org/10.1016/0885-2308(91)90019-M.

[73] C. Sidner, C. Lee, and N. Lesh, "Engagement when looking: Behaviors for robots when collaborating with people," Dec. 2003.

[74] P. J. Doyle, K. de Medeiros, and P. A. Saunders, "Nested social groups within the social environment of a dementia care assisted living setting," *Dementia*, vol. 11, no. 3, pp. 383–399, 2012. DOI: 10.1177/1471301211421188.

[75] M. Campo and H. Chaudhury, "Informal social interaction among residents with dementia in special care units: Exploring the role of the physical and social environments," *Dementia*, vol. 11, pp. 401–423, May 2012. DOI: 10.1177/1471301211421189.

[76] M. P. Calkins, "From research to application: Supportive and therapeutic environments for people living with dementia," *The Gerontologist*, vol. 58, S114–S128, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:26108257.

[77] J. Cohen-Mansfield, T. Hai, and M. Comishen, "Group engagement in persons with dementia: The concept and its measurement," *Psychiatry Research*, vol. 251, Feb. 2017. DOI: 10.1016/j.psychres.2017.02.013.

[78] S. Shen, P. Slovak, and M. F. Jung, "Stop. I see a conflict happening.: A robot mediator for young children's interpersonal conflict resolution," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18, Chicago, IL, USA: Association for Computing Machinery, 2018, pp. 69–77, ISBN: 9781450349536. DOI: 10.1145/3171221.3171248.

[79] M. F. Jung, N. Martelaro, and P. J. Hinds, "Using robots to moderate team conflict: The case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15, Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 229–236, ISBN: 9781450328838. DOI: 10.1145/2696454.2696460.

[80] E. Short and M. J. Mataric, "Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 385–390. DOI: 10.1109/ROMAN.2017.8172331.

[81] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Šabanović, *Human-robot interaction: An introduction*. Cambridge University Press, 2020.

[82] C. U. Krägeloh, J. Bharatharaj, S. K. Sasthan Kutty, P. R. Nirmala, and L. Huang, "Questionnaires to measure acceptability of social robots: A critical review," *Robotics*, vol. 8, no. 4, 2019, ISSN: 2218-6581. DOI: 10.3390/robotics8040088.

[83] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies*, vol. 7, pp. 437–454, Nov. 2006. DOI: 10.1075/is.7.3.14nom.

[84] C. Carpinella, A. Wyman, M. Perez, and S. Stroessner, "The robotic social attributes scale (rosas): Development and validation," Mar. 2017, pp. 254–262. DOI: 10.1145/2909824.3020208.

[85] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, pp. 71–81, Jan. 2008. DOI: 10.1007/s12369-008-0001-3.

[86] A. Peca, M. Coeckelbergh, R. Simut, *et al.*, "Robot enhanced therapy for children with autism disorders: Measuring ethical acceptability," *IEEE Technology and Society Magazine*, vol. 35, no. 2, pp. 54–66, 2016. DOI: 10.1109/MTS.2016.2554701.

[87] P. Alves-Oliveira, T. Ribeiro, S. Petisca, E. Di Tullio, F. S. Melo, and A. Paiva, "An empathic robotic tutor for school classrooms: Considering expectation and satisfaction of children as end-users," in *International Conference on Social Robotics*, Springer International Publishing, 2015, pp. 21–30.

[88] D. S. Syrdal, T. Nomura, and K. Dautenhahn, "The frankenstein syndrome questionnaire – results from a quantitative cross-cultural survey," vol. 8239, Oct. 2013, pp. 270–279, ISBN: 978-3-319-02674-9. DOI: 10.1007/978-3-319-02675-6_27.

[89] T. Ninomiya, A. Fujita, D. Suzuki, and H. Umemuro, "Development of the multi-dimensional robot attitude scale: Constructs of people's attitudes towards domestic robots," vol. 9388, Oct. 2015, pp. 482–491, ISBN: 978-3-319-25553-8. DOI: 10.1007/978-3-319-25554-5_48.

[90] L. Veling and C. McGinn, "Qualitative research in HRI: A review and taxonomy," *International Journal of Social Robotics*, vol. 13, no. 7, pp. 1689–1709, Nov. 2021.

[91] M. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: A review," *Smart Learning Environments*, vol. 6, Jan. 2019. DOI: 10.1186/s40561-018-0080-z.

[92] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: Common objects in context," vol. 8693, Apr. 2014, ISBN: 978-3-319-10601-4. DOI: 10.1007/978-3-319-10602-1_48.

[93] F. Robinson and G. Nejat, “An analysis of design recommendations for socially assistive robot helpers for effective human-robot interactions in senior care,” *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 9, p. 20 556 683 221 101 389, 2022. DOI: 10.1177/20556683221101389.

[94] M. L. Walters, K. L. Koay, D. S. Syrdal, K. Dautenhahn, and R. te Boekhorst, “Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials,” 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:8049157.

[95] B. Zhang, J. Amirian, H. Eberle, J. Pettré, C. Holloway, and T. Carlson, “From HRI to CRI: Crowd robot interaction - understanding the effect of robots on crowd motion,” *International Journal of Social Robotics*, vol. 14, no. 3, pp. 631–643, Apr. 2022, ISSN: 1875-4805. DOI: 10.1007/s12369-021-00812-7.