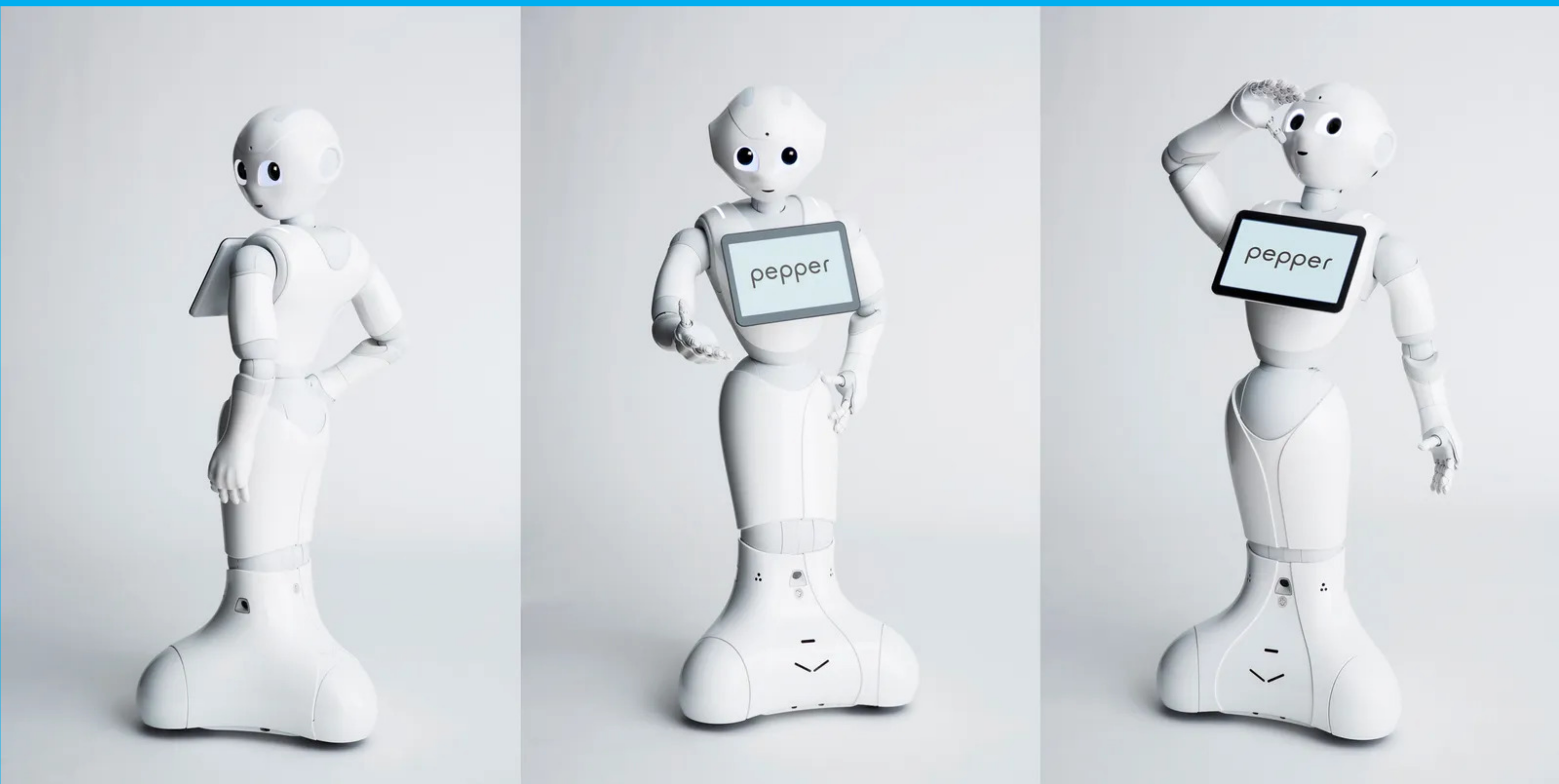
# Human- or Robot-like Music Assistive Robots

## Effects on Fluency and Memory Recall

Yanzhe Li




TUDelft

# Human- or Robot-like Music Assistive Robots

## Effects on Fluency and Memory Recall

by

Yanzhe Li

| | | |
|---|---|---|
| Student number: | 5220033 | |
| Project duration: | October 1, 2021 – September 26, 2022 | |
| Thesis committee: | Prof. dr. M.A. Neerincx, | TU Delft, supervisor |
| | Dr. F. Broz, | TU Delft |
| | Dr. U. Gadiraju, | TU Delft |

TU Delft

# Abstract

Longer lifespans and an ageing population put tremendous pressure on the care of the elderly. With the technology of robotics breakthroughs, it appears that the use of robotics in elderly care is ready to take off. Interestingly, more and more robots are now being created with a human-like appearance and demeanour, yet this could lead to issues like inappropriate mental models of robots in the user's mind. On the other hand, due to their minimal communication challenges, robot-like robots may function effectively in the human-robot interaction. Up to now, far too little attention has been paid to the increasing trend of the human-like robot. This thesis compared the performance of these two types of robots, which may help to explain why the trend is the way it is. Two research questions are created to accomplish this goal: Which type of robot, robot-like or human-like, produces the most fluent interaction in the elderly care scenario? Does the robot-like or human-like robot bring about the best memory performance in the music listening activity?

This thesis project aims to contribute to the development of the social robot in elderly care. In this thesis, a music assistance robot was developed, it can converse with the user while listening to music and recalling memories associated with the song. The elder's health and well being are benefited by this pastime. To address the question of which type of robot performs better, two types of robots, a human-like robot and a robot-like robot, were created to compare in terms of fluency in human-robot interactions and performance of memory recall during music listening activities. Four key elements: appearance, behaviour, voice and dialogue were used in the construction of two robot identities. Based on the robot we created, a joint music listening experiment was carried out. For evaluation, a questionnaire and video recordings of the interaction were both employed. This was tested with 30 volunteers (The young adults were chosen for the experiment due to COVID restrictions and time limitation), and it was discovered that the human-like robot performed better overall.

**Keywords"** Robot; Music; Human-like robot; Robot-like robot; Fluency; Memory recall

# Contents

# List of Figures

# List of Tables

# 1 Introduction

People are living longer and the population is ageing. The number of people aged 60 years and older will rise to 2.1 billion by 2050, the population is ageing at a considerably higher rate than in the past (WHO, 2021). According to the United Nations (2020), there are 727 million people over the age of 65 in 2020. Based on these facts, the importance of elderly care is widely recognized (Hannah Richardson, 2017; Pfadenhauer & Dukat, 2015). The rise in the number of elderly people is putting tremendous pressure on elderly care providers.

With the numerous advancements in robotics, the application of robotics in elderly care seems ripe for development. These robots, which can be human-like or robot-like, could be used for a variety of purposes, including healthcare, general help, artificial companions, and leisure activities (Sharma et al., 2021). Both types of robots have their own benefits, the human-like robot is more similar to humans in terms of social interaction (Duffy, 2003), thus, humans are able to communicate more naturally with the robot in social scenarios. However, as a robot's human-likeness improves, so do the expectations for understanding and intelligent responses (Rothstein et al., 2020), which may not always match actual robot capabilities. Especially the elderly people are often unfamiliar with the cutting-edge technology, they may anticipate human-like robots to be able to behave like a real human that are now impossible to achieve with current technology. On the contrary, if robots act in a robot-like manner, people will not expect they have a high level of social cognition. In this way, the elderly are more likely to develop appropriate expectations and trust. In brief, people's mental models of robots may be inaccurate or overconfident, which may result in inappropriate interactions with robots. Darling (2021) claims that it would be better for humans if we treated robots with a bit of humanity, more in line with how we treat animals. In fact, both types of robots are utilised in various elderly care scenarios, and there are an increasing number of human-like robots in robotics applications (Erebak & Turgut, 2019), However, because elderly people may have inappropriate mental models of the robots, it can be difficult for robots to create "fluency" interactions with them. Additionally, a robot-like robot may not have the same fluency as a human-like robot.

The care of older people is a major concern in today's society, and there are numerous initiatives that focus on providing entertainment, encouraging mobility, and increasing socialising, nevertheless the most prevalent are those aimed at improving health and well-being, with music is one of the things that often involved here (Bandini et al., 2019). Especially, studies demonstrated that music can effectively trigger autobiographical memories (Belfi et al., 2016; Janata et al., 2007), and as noted by Hays and Minichiello (2005), it can help the elderly by connecting them to "others who may no longer be living, and may also validate memories, give meaning to live, and bring a greater sense of spirituality". In conclusion, music-evoked autobiographical memories (MEAMs) is also an area of concern in elderly care.

In this study, we would like to develop a social robot that supports elderly in several daily activities, such as listening to music together and sharing music-related experiences. Music has been shown to enrich the health and well-being of the elderly. Therefore, we are aiming at the combination of a social robot and music. It can, however, be created as either a human-like or a robot-like robot, and very little studies have looked into the implications. As a result, our research questions are:

- **Which type of robot, robot-like or human-like, produces the most fluent interaction in the elderly care scenario?**
- **Does the robot-like or human-like robot bring about the best memory performance (MEAMs) in the music listening activity?**

Unfortunately, these two research questions were initially tested with young adults rather than vulnerable older persons for two reasons: First, we aim to put as little strain as possible on the target group with our research and first gain a general understanding of the relevant phenomena. A second practical reason is that it takes up a significant amount of the limited time to recruit older people.

## 1.1. Thesis outline

The literature review for this thesis is covered in the following chapter. Chapter 3 contains the methods utilized to address the research questions. The implementation of the experiment are described in Chapter 4. Chapter 5 presents the experiment's results. Chapter 6 discuss the results and the conclusions that are drawn from them, future work is also covered here.

# 2

# Literature Review

## 2.1. Fluency of human-robot interaction

Over the past few decades, there has been a significant increase in using socially capable robots. In fact, Interaction between humans and robots has become a necessity. To improve the quality of human-robot interaction, practitioners and researchers started to focus on the concept of fluency.

Humans are capable of reaching a high level of coordination while working together on a common task, which results in a well-synchronized meshing of their activities. This conduct frequently manifests itself without the exchange of many verbal cues because of their exact and effective timing, dynamic adjustment of their plans and actions. This quality of interaction is the definition of the *fluency* of the shared activity (Hoffman, 2019). To date, several studies have investigated the fluency in HRI research. Chao and Thomaz (2012) examined fluency to assess a multimodal turn-taking system, Cakmak et al. (2011) evaluated the fluency of handovers from a robot to a human. However, assessing it is challenging due to the lack of uniform criteria. Research on this subject has been focused on determining the fluency of human-robot interaction in a collaborative work setting (e.g., shared workspace tasks, handover tasks, shared manipulation tasks). In this study, we would like to use it for the joint music listening activity, as we think that this concept is also relevant for such (joint) cognitive activities. To the best of our knowledge, the fluency measurement of human-robot interaction can be broadly divided into two approaches: objective and subjective measures.

Most studies of objective fluency metrics have only been carried out in measuring the time interval during the human-robot interaction activity (Baraglia et al., 2016; Huang et al., 2015; Isaacson et al., 2020; Maniadakis et al., 2017; Nikolaidis & Shah, n.d.; Rahman, 2018). In most circumstances, the metrics can be listed as follows: the percentage of concurrent activity (C-ACT), the human's idle time (H-IDLE), the robot's functional delay (F-DEL), and the robot's idle time (R-IDLE). (see Table 2.1)

Subjective fluency metrics, on the other hand, can be used to assess how fluent humans perceive human-robot interactions. Much of the previous research on these metrics have used questionnaires to evaluate agreement with the concept of fluency. Although subjective sensations can be elicited through a number of different approaches, the majority of them fall into the following categories: robot trust, robot contribution, and robot positive teammate traits (Dragan Anca D. et al., 2015; Hoffman & Breazeal, 2007; Unhelkar et al., 2014). Hoffman (2019) summarized the previous studies and suggested a composite measure, which includes 7 metrics, to evaluate the subjective fluency perception. However, the "improvement" metric only applies to learning and adaptation scenarios, hence, it falls out of the scope of this study and will not be covered here. All of the other 6 metrics listed in Table 2.2 can be used in this thesis.

In this thesis, a combination of objective and subjective metrics was used. For the objective metrics, Hoffman (2019) points out that only F-DEL and H-IDLE show a significant correlation with fluency perception. However, F-DEL is for multi-agent settings, which don't apply in our situation, hence we chose H-IDLE as the objective metric. For the subjective metrics, researchers discovered that overall

| Metrics | Description |
|---|---|
| H-IDLE | It is the percentage of total task time when the human is inactive. In most cases, the human is waiting for the robot to finish actions since the humans have faster perceptual processing. |
| R-IDLE | It is the percentage of total task time when the robot is inactive. R-IDLE can happen in a variety of situations: e.g., waiting for human input; process input; decision making; waiting for a human to finish an action. |
| C-ACT | It is the percentage of total task time when both the human and robot are active. Typically, this metric is used to measure the collaborative (joint) and handover tasks in a shared workspace, i.e. the tasks the human and the robot can perform simultaneously. |
| F-DEL | It is the percentage of total task time between the end of one agent's action and the start of the other agent's action expressed. The overall F-DEL is roughly equivalent to that imposed by the robot because, in reality, human F-DEL is frequently insignificant. |

Table 2.1: Objective fluency metrics. Retrieve from Hoffman, 2019

fluency, trust, and robot contribution were the most effective subjective indicators, which is consistent with Hoffman's (2019) findings. As a result, we employed the first three metrics in Table 2.2 in this thesis.

## 2.2. Mental model in human-robot interaction

Wickens et al. (2015) defined a Mental Model as a mental structure that reflects understanding of a system by the user, and how they expect the system to respond. In the field of human-robot interaction, this concept refers to the user's expectations of the robot's behaviour or capabilities. It is now well established from a variety of studies that people's mental model of robots can be influenced by numerous aspects of the robot, e.g., appearance (Kwon et al., 2016), behaviours (Goetz et al., 2003), and social cues (Kiesler, 2005). However, people's expectations (i.e., humanoid robots can carry on conversations) may not always match actual robot capabilities (Schramm et al., 2020). Furthermore, when mixed with confirmation bias (Nickerson, 1998), individuals frequently have inaccurate mental models of robots that are far from their actual capabilities (Haring et al., 2018; Paepcke & Takayama, 2010). As a result, interaction failures may occur (Preeti Ramaraj, 2021), and the user may even refuse to use the system (Graaf et al., 2017). To date, several studies have investigated how to construct a more effective mental model to improve human robot interactions (Lee et al., 2005; Powers & Kiesler, 2006). To determine the effects of robot design, Frijns and Schürer (2020) and Kiesler and Goetz (2002) compared the impact of human-like and machine-like robots on people's mental models. In the above-mentioned papers, questionnaires are often used to measure the mental models. In this thesis, these significant studies are utilized to create appropriate participants' mental models of robots that used in the experiments. More specifically, before to the experiment, participants will be given an introduction to assist them in developing an appropriate mental model (Ligthart et al., 2017).

## 2.3. Music-evoked autobiographical memories

Birthday celebrations, weddings and funerals are all occasions where music is played and these music compositions are inextricably linked to these life events. Hearing these musical pieces can recall autobiographical memories and the feelings linked with them, even after a considerable time has passed. Schulkind et al. (1999) reported that there was a significant positive correlation between emotion and memory, particularly for older persons. Due to the fact that strong emotions enhanced memory processes and music induces strong emotions, music may have a role in memory formation (Jäncke, 2008). There is a large number of published studies (e.g., Belfi et al., 2016; Jakubowski and Ghosh, 2021; Janata et al., 2007; Platz et al., 2015) that look into the music-evoked autobiographical

| **Metrics** | **Description & Indicators** |
|---|---|
| Human-robot Fluency | Evaluate the overall fluency between the human and the robot.<br>• "The human-robot team worked fluently together."<br>• "The human-robot team's fluency improved over time."<br>• "The robot contributed to the fluency of the interaction." |
| Robot Contribution | Evaluate the robot's contribution to the team<br>• "I had to carry the weight to make the human-robot team better." (reverse scale)<br>• "The robot contributed equally to the team performance."<br>• "I was the most important team member on the team." (reverse scale)<br>• "The robot was the most important team member on the team." |
| Trust in Robot | Evaluates the trust the robot evokes.<br>• "I trusted the robot to do the right thing at the right time."<br>• "The robot was trustworthy." |
| Robot Teammate Traits | Evaluate the robot's perceived character traits in relation to its role as a team member.<br>• "The robot was intelligent."<br>• "The robot was trustworthy."<br>• "The robot was committed to the task." |
| Working Alliance for Human-Robot Teams | Evaluate the quality of working alliance of human-robot teamwork, and consists of two subscales: "bond" sub-scale and "goal" sub-scale.<br>**"Bond" sub-scale**<br>• "I feel uncomfortable with the robot." (reverse scale)<br>• "The robot and I understand each other."<br>• "I believe the robot likes me."<br>• "The robot and I respect each other."<br>• "I am confident in the robot's ability to help me."<br>• "I feel that the robot appreciates me."<br>• "The robot and I trust each other."<br>**"Goal" sub-scale**<br>• "The robot perceives accurately what my goals are."<br>• "The robot does not understand what I am trying to accomplish."(reverse scale)<br>• "The robot and I are working towards mutually agreed upon goals." |
| Individual Measures | Additional useful indicators.<br>• "The robot's performance was an important contribution to the success of the team."<br>• "It felt like the robot was committed to the success of the team."<br>• "I was committed to the success of the team." |

Table 2.2: Subjective fluency metrics. Retrieve from Hoffman, 2019

memories (MEAMs). These studies show that music can stimulate a broad variety of memories, with one notably finding is that the memory of elderly people in the "reminiscence bump" period is disproportionate compared to other lifetime periods. The reminiscence bump is the tendency for older adults recall experiences from their adolescence and early adulthood (10-30 years of age) with augmented or enhanced recollection (Munawar et al., 2018). In addition, music has been demonstrated to improve autobiographical memory in dementia patients (El Haj, Postal, et al., 2012).

For measurement, a reliable instruction is known as the TEMPau test (Test Episodique de Mémoire du Passé) is widely used to assess autobiographical memory performance (El Haj, Fasotti, et al., 2012; El Haj et al., 2018; El Haj et al., 2016; Piolino et al., 2006). To eliminate scoring bias, most of these studies used a second independent rater to score the data (20% of the data). Participants were also given up to five minutes to describe their memories to minimise repetitive memory or distractibility.

The TEMPau test ranges from 0 to 4 points, 0 for no memory or only general information; 1 for memory in a repeated/extended event; 2 points is the same as 1 point but with spatiotemporal details; 3 points if the memory was specific with spatiotemporal details; and 4 points will be given if it with the presence of internal sensory-perceptual-affective details on the condition of 3 points. (See example in Table 2.3)

| Score | Example |
| --- | --- |
| 0 | My father. |
| 1 | My father used to drink coffee. |
| 2 | My father used to drink coffee in the backyard. |
| 3 | One morning on a summer vacation in the mountain, my father was not able to find a grocery to buy coffee. |
| 4 | One morning on a summer vacation in the mountain, My father was bit nervous without his morning coffee. |

Table 2.3: TEMPau test score example from El Haj et al., 2018

## 2.4. Music and elderly care

Music's significance in health promotion has been studied throughout history. Music was utilized to boost spirits, fend off illness, and give an overall relaxing effect in ancient societies of Asia, Egypt, Romania, Africa, and the Americas (Mileski et al., 2019). Nowadays, people's life expectancy is getting longer and longer, and the elderly population is steadily growing, as is the demand for elder care. While in the elderly care facilities, music has been demonstrated to be a valuable resource for reliving agitation, stress, and depression for many years (Fu et al., 2018; Gupta et al., 2021; Lou & Mn, 2001). Particularly, music has been shown to promote health and well-being in the elderly, it can be done in a variety of ways, including singing (Skingley & Bungay, 2013), listening (Wattanasoei et al., 2017) and dancing (Douka et al., 2019). Moreover, music can be used in reminiscence therapy for the elderly, which is commonly used in dementia care and has been demonstrated to improve the well-being of the patient (Baird & Thompson, 2018; Moreno-Morales et al., 2020). Especially, more recent attention focused on combining the social robot and music in dementia care (Kok et al., 2018; Peeters et al., 2016).

## 2.5. Big 5 personality in human-robot interaction research

Personality is defined as a person's behaviours, cognition, and emotions, which are influenced by both biological and social factors (Hall & Lindzey, 1957). Social psychology research has demonstrated that people with different personalities prefer to interact in different ways (Cruz-Maya & Tapus, 2016). Personality is increasingly recognised as a key concept in understanding human behaviour (Goyal et al., 2008; Li et al., 2014).

The Big 5 personality traits are commonly employed in research (Li et al., 2014), which are made up of Conscientiousness, Agreeableness, Emotional Stability, Openness to Experience, and Extraversion. Significantly, the Big 5 personality traits were often included in the human-robot interaction study (Damholdt et al., 2015; Salem et al., 2015; Sandoval et al., 2016; Walters, Syrdal, Dautenhahn, et al., 2008). For instance, Salem et al. (2015)'s research revealed that people with higher emotional stability have greater psychological connection and likability toward human-like robots. Moreover, Walters, Syrdal, Dautenhahn, et al. (2008) discovered that persons with higher emotional stability preferred human-like robots over robot-like robots. Furthermore, another study (Gockley & MatariĆ, 2006) on how robots encourage humans to exercise discovered no significant impacts related with any of the Big 5 personalities.

All in all, the Big 5 personality can help researchers in the HRI area gain a better understanding of the performance of human-robot interaction. However, to the best of our knowledge, no related research has been conducted on the effects of personality on fluency and memory recall in a music HRI scenario. As a result, the Big 5 personality traits will be investigated in this study as well.

## 2.6. Conversation between human and computer

To the best of our knowledge, there are no explicit design principles to distinguish between human-like and robot-like robots for the dialogue of human-robot interaction. However, Fischer (2006) has summarised the existing literature on the differences between human-to-human communication (HHC) and human-computer interaction (HCI) in various aspects (see Table 2.4).

| **Aspects** | **Description** |
|---|---|
| Politeness: | Morel (1989) points out that while communicating with artificial systems, the speakers are less polite. However, some analysts (e.g., Richards and Underwood (1984) and Ward and Heeman (2000)) use a variety of sources to determine that speakers in HCI are just as polite as in HHC circumstances if the system is also polite. |
| Verbosity: | On this aspect, researchers have a variety of viewpoints. Amalberti et al. (1993) discovered that HCI speakers produced much more words each dialogue than HHC, but Oviatt et al. (2004) discovered even less. |
| Variability: | Much of the literature since the 1990s (Amalberti et al., 1993; Gustafson, 2002) emphasises that HCI has a smaller lexicon and fewer syntactic structures. |
| Simplification: | Gustafson (2002) suggests that HCI has a lot of simplification, such as a lack of determiners, conjunctions, and prepositions, as well as a basic and limited syntax. |
| Overspecification: | Fischer (2006) have found that the amount of overspecifications (e.g. prepositions, additional determiners, conjunctions) increases in HCI when compared to HHC. However, Hitzenberger and Womser-Hacker (1995) was unable to statistically confirm this theory. |
| Discourse structure: | Early examples of research into discourse structure show that there are fewer connectives in HCI (Amalberti et al., 1993; Krause, 1992). Oviatt et al. (2004) attributes this to the lack of structure in human-computer interaction. However, other authors (Amalberti et al., 1993; Porzel & Baudis, 2004) argued that the user could maintain to employ structuring devices or even repeat the grounding utterance in HCI. |

Table 2.4: The differences between human-to-human communication and human-computer interaction

# 3 Methodology

As explained in the introduction, Chapter 1, this thesis contains two research questions and this experiment aims to validate these research questions. The experiment was primarily focused on two dimensions: fluency and the performance of MEAMs. To evaluate these two metrics, a user study was conducted. Participants were divided into two groups, one of which was engaged with a human-like robot and the other with a robot-like robot. The robot is controlled by an experimenter behind the curtain, which is known as the Wizard of Oz experimental method (Dahlbäck et al., 1993). The entire interaction, including audio and video, was recorded by cameras. Based on the data collected by the camera, the objective metrics of fluency and performance of MEAMs can be determined. Following the completion of the experiment, participants was given a questionnaire in which they were asked about their subjective feelings of fluency when interacting with the robot. In addition, the Big 5 personality traits were examined. Finally, through the analysis of objective and subjective data, research questions can be answered.

## 3.1. Participants

The participants were students from Delft University of Technology. About half of them were recruited by social media advertisements, and the other half were recruited from the Computer Science building on the day of the experiment.

The experiment involved 33 participants in total (N = 33). Three of the video recordings are, however, absent due to technical issues. As a result, 30 participants in total (N=30, 91%) were available for data analysis. The participants were evenly divided into two groups for these 30 valid data, with 15 being assigned to the robot-like robot condition and the other 15 to the human-like robot condition. The average age was 24.8 years old (SD = 4.8).

## 3.2. Material

The website Qualtrics [1] was used to create the questionnaire and the informed consent form. A tablet was used for the participants to fill in these forms. The human-robot interaction was recorded using a GoPro camera mounted on a tripod. In the experiment, a Pepper robot was employed. An interface that had been modified from the work of WoZ4U (Rietz et al., 2021) was used to control the robot, it was thoroughly covered in Section 4.1.2. The experiment's music was played via a Bluetooth music box that was connected to the laptop. A laptop and an extra monitor were used to control the experiment. Some candies and snacks were provided during the experiment.

## 3.3. Experimental design

As mentioned in Chapter 1, we formulated two research questions in this study. Table 3.1 lists the research question's variables and demonstrates that the independent variable for both research

[1] https://www.qualtrics.com

questions is the robot's identity. The fluency of the human-robot interaction serves as the dependent variable for the first research question. It is the performance of the memory recall for the second research question. The measures were described in Section 3.6 in detail.

| **Research question** | **Independent variable** | **Dependent variable** | **Measures** |
|---|---|---|---|
| Which type of robot, robot-like or human-like, produces the most fluent interaction in the elderly care scenario? | The robot identity | Fluency of the human-robot interaction | H-IDLE ratio (objective) and questionnaire (subjective) |
| Does the robot-like or human-like robot bring about the best memory performance (MEAMs) in the music listening activity? | The robot identity | Performance of the memory recall | TEMPau test score |

Table 3.1: The independent and dependent variables

Two hypotheses were proposed according to these research questions (see Table 3.2). To test these hypotheses, two types of robot were designed. As mentioned in Chapter 1, they have the same function but with different identities, one is a human-like robot and another one is a robot-like robot.

| | **Hypothesis ($H_1$)** |
|---|---|
| Fluency | The robot-like robot can produces the most fluent interaction in the music listening activity. |
| MEAMs | The human-like robot bring about the best memory performance in the music listening activity. |

Table 3.2: The hypothesis

## 3.4. Procedure

We first applied for ethical approval for this study from the Delft University of Technology's Human Research Ethics Committee. After approval was obtained, we started recruiting participants for the experiment. The participants were asked to come to the INSYGHTLab on the Delft campus to participate in the experiment and to sign consent forms prior to doing so. The set-up of the experiment can be seen in Figure 3.1.

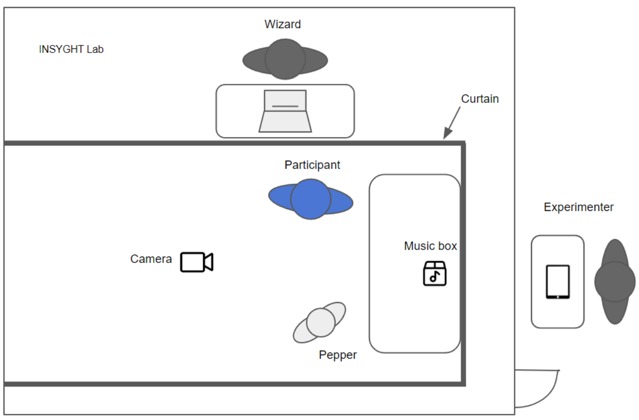



Figure 3.1: Schematic set-up of the experiment

It took roughly 15 minutes to complete the experiment. The experimenter began by giving a brief introduction and asking the participant to sign the informed consent. Then they were instructed to take a seat on the chair (see Figure 3.2) and informed that the experiment would shortly begin. The experimenter would now activate the robot and launch the experiment, Figure 3.3 shows the main procedure of the experiment.The robot started with an introduction and then asked the participants for their profile information. Once the profile is created, the robot helped the participant choose a song that they know. Once the selection was made, the song was played and the robot started dancing. At the same time, the robot asked general questions about the music. When the dancing stopped, the robot asked for some memories related to the song. The experimenter then checked the TEMPau score of the participant's speech and if it is below 2, the experimenter tried to ask for more information about the memory. If the participant has already been asked twice about memory, the robot will end the conversation. That is the main procedure of the experiment, the participants were given a questionnaire once it was completed.

Figure 3.2: A site photos of the experiment

## 3.5. Fail-safe protocols

Because this is a human-related study, the participants may not completely follow our scripts for doing the experiment, several fail-safe protocols were used to avert unforeseen difficulties during the experiment. Firstly, some generic answers like ”um”, ”yes, that's correct” are embedded in the interface. Secondly, there is an input box which can type and directly send the messages to the robot, with these two approaches, most unforeseen problems can be solved (see Figure 4.2). Thirdly, if the experiment becomes uncontrollable, the experimenter can immediately stop it. Finally, if the experiment becomes uncontrollable or the participant wants to quit, the experimenter can stop it immediately.

## 3.6. Measures

During the experiment, two types of data were collected: video records of the experiment and the follow-up questionnaires. We already discussed the data we intend to collect in the Chapter 2. In this section, the data was detailed discussed as well as how we handle the data.

### 3.6.1. Details of the data

- Two study variables are included in the video observations section: the objective fluency metrics and TEMPau test score. The objective fluency metrics can reflect how well both experiment conditions performed in terms of fluency. The TEMPau test score can show how well the participants performed memory recall, it is determined by the participant's speech during the

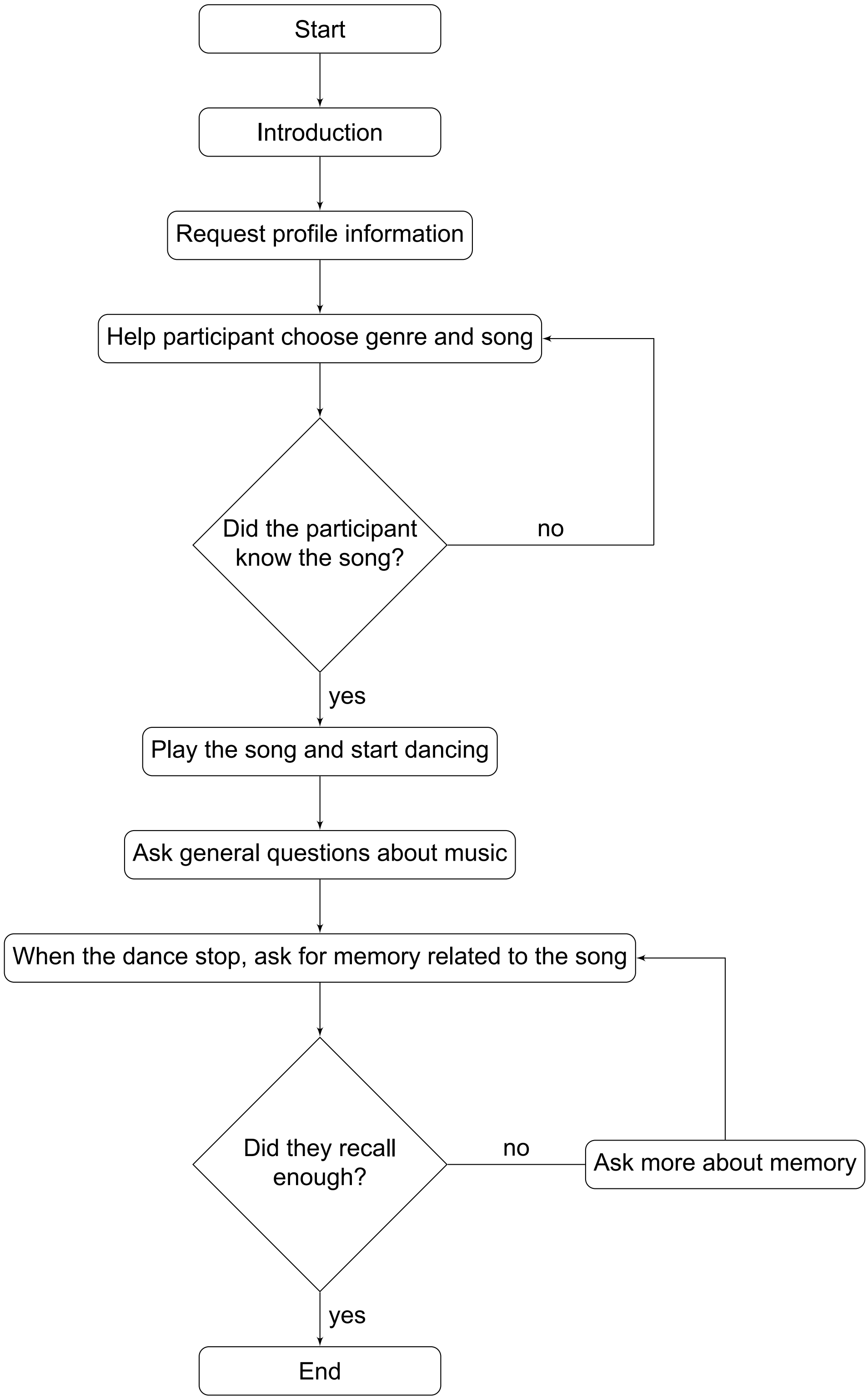



Figure 3.3: The procedure of the experiment

memory recall session.

- The questionnaire is divided into two main sections, the first of which discusses the subjective fluency metrics and the second of which discusses the Big 5 personality test. The questionnaire's questions were in the appendix A. Subjective fluency metrics contain nine questions, all of which used a 7-point Likert scale. Ten-Item Personality Inventory (TIPI) scoring scale (Gosling et al., 2003) was used to assess the individuals' Big 5 personality traits.

### 3.6.2. How to handle the data

The objective fluency metric was measured by the H-IDLE time ratio of the video observation, it can be directly measured by watching the video records and counting the time ratio of the human is inactive. To achieve this goal, a video player and a stopwatch were used. Play the recorded video in the player first, then start the stopwatch when the participant was engaged (e.g., talking or touching the tablet) and stop it when they weren't (e.g. robot starts talking). Finally, divide the stopwatch's time by the length of the interaction that preceded the memory recall session, which was the H-IDLE.

For the TEMPau test score, it was measured based on the scale from Table 2.3, and a second independent rater assessed 20% of it.

## 3.7. Data analysis

The data was collected from the video camera and the post-experiment questionnaire. In general, the evaluations would follow the guidelines in Sections 2.1, 5.4 and 2.3 of the literature review. Following is a list of these three factors.

- **Fluency**: A combination of objective and subjective metrics was used. **H-IDLE** was utilised for objective metrics, and **overall fluency**, **trust**, and **robot contribution** were used for subjective metrics.
- **MEAMs**: The **TEMPau test** was used to measure the performance of the MEAMs.
- **Big 5 Personality traits**: In addition, a big 5 personality traits (TIPI) were measured for every participant in the questionnaire.

Fluency and MEAMs were two of the characteristics utilised to compare the different types of robot identities, and we performed hypothesis testing to determine whether there was a significant difference. A normality test was conducted first due to the small sample size. The t-test was used if the normality test was passed by both datasets, otherwise the Mann-Whitney U test was employed.

In this study, the relationships between the variables (fluency metrics and the TEMPau test score) and the big 5 personality traits were examined by administering Pearson correlation test.

# 4

# Experimental Implementation

## 4.1. Technology

### 4.1.1. Pepper robot functionalities

The humanoid Pepper robot (Figure 4.1 [1]) is developed by Softbank Robotics (Softbank, 2022). Pepper has 20 degrees of freedom for natural and expressive movements, along with touch sensors, LEDs and microphones for multimodal interactions. Apart from that, Pepper has a variety of advanced sensors, including Infrared sensors, an inertial unit, 2D and 3D cameras, sonars and many other sensors. It measures 120 centimeters tall and weighs 28 kilograms. Pepper is one of the most widely used research platforms in HRI, and it has also been utilised in real-world applications such as greeting customers in businesses (Aaltonen et al., 2017) and guiding visitors through museums (Allegra et al., 2018). In the evaluation studies, this robot was employed.

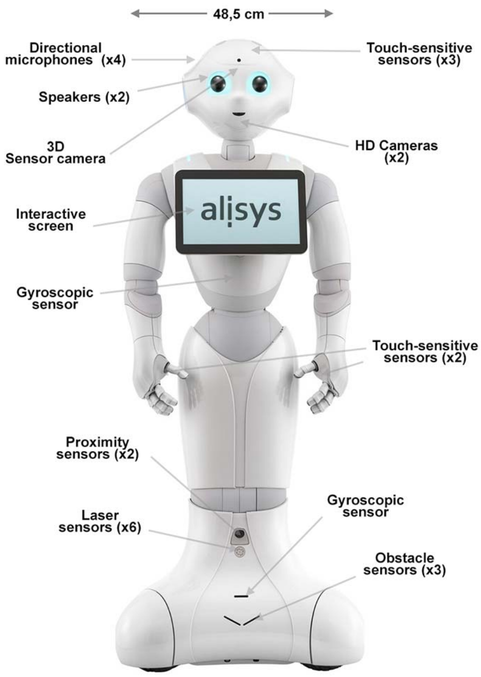



Figure 4.1: An image of Pepper

[1]Retrieved from https://tecnoemergentes.foroactivo.com/t152-pepper-el-robot-cognitivo

### 4.1.2. Robot software platform

The control method of the robot was developed by using the NAOqi SDK [2] and Choregraphe [3], which are programmed in Python. These two tools, provided by Softbank Robotics, are widely used by researchers (Pot et al., 2009). Choregraphe is a graphical interface to program the Pepper in a drag-and-drop manner. It offers various advantages, including being intuitive, easy to use, and friendly to users with no programming knowledge. The integrated virtual robot, in particular, simplifies the creation of robot motion. However, like with any software tool, lower programming skills necessitate a reduction in versatility, when the experimental design demands the robot to respond fast and adapt to the participant's behaviour, Choregraphe is not a good fit. To gain all the functionality of the Pepper robot, the NAOqi Python SDK was used to develop the Wizard-Of-Oz tool. Since the Pepper was widely used in the HRI research, Rietz et al. (2021) developed an open-source Wizard-Of-Oz interface (WoZ4U [4]) for the Pepper. Nevertheless, the majority of this tool's functions were useless for this experiment. We modified this tool base on its framework and kept 4 functions, which were the connection method to the robot; message to speech; voice setting and the autonomous life setting. All these functions can be found in Figure 4.2, it is worth noting that the messages in the "main dialogue" can also perform robot gestures (e.g. greetings). For the other parts, they stick to the original intent. To display the music playlist on the tablet and perform the dance, a new area called "Playlist & Dance" was added to this interface. The dance can be triggered by simply clicking the button and stopped by clicking it again, as well as the display music playlist.

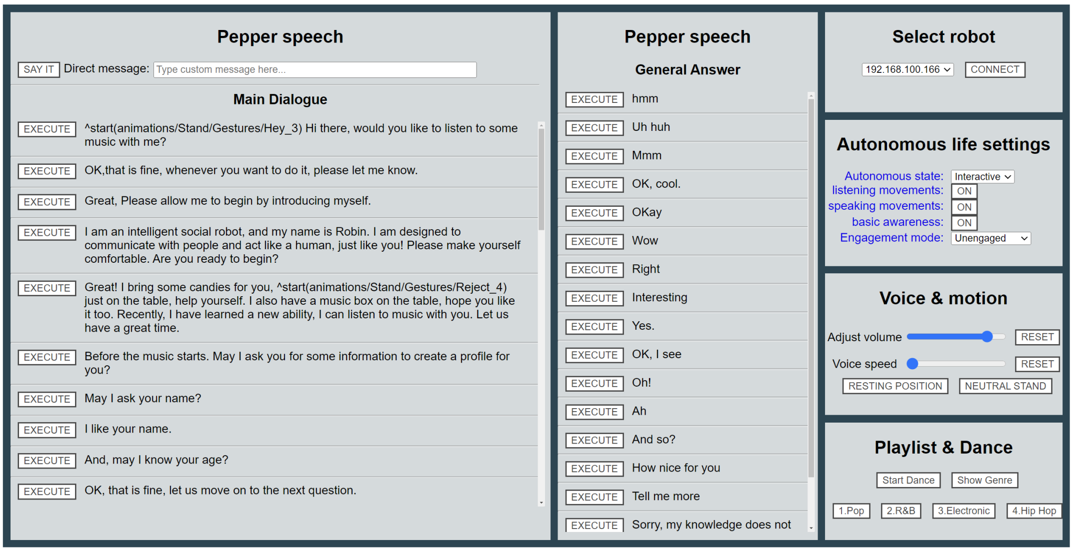



Figure 4.2: The Interface of the modified WoZ4U

## 4.2. Design of different robot conditions

The design of different robot conditions relies on four main aspects: appearance, behaviour, voice, and dialogue.

- **Appearance** The user's first impression of the robot is its appearance. Therefore, the appearance of the robot is crucial in helping the user to develop the correct mental model with the robot. In this study, the robot-like type robot maintains its original appearance, whereas the human-like type robot dresses in clothing with noticeable "human" traits, such as a tie and a hat.

- **Behaviour** The user's perception of the robot's identity during the interaction is heavily influenced by the robot's behaviour, such as the robot's body movements when talking to the user. However, a number of authors have proposed that the impact of gaze and gestures are essential communication aspects are necessary to be included in HRI (Kompatsiari et al., 2017;

[2]http://doc.aldebaran.com/2-8/index_dev_guide.html
[3]http://doc.aldebaran.com/2-4/software/choregraphe/index.html
[4]https://github.com/frietz58/WoZ4U

Salem et al., 2011). Likewise, Fischer (2019) suggested that the collaborative robots should be social actors, which necessitates the usage of social cues and may involve the display of emotions. Therefore, all gestures are kept the same with the exception of the dance, which is done for the reason that it won't affect the essential communication between the human and the robot.

Abe et al. (2020) have summarized that a human-like robot dance should have the following characteristics: human quality motion, flowing, organic, natural, curved lines, whereas the robot-like robot dance should have: precision, control, proximity and safety. In this study, the dances of both robots were designed following the guidelines of these principles. For the human-like dance, some obviously human-like actions were added to the human-like dance to help the participants form a human-like perception of the robot. For example, the robot extend its hands as an invitation at the start of the dance, then place one hand on the ear and wave the other hand in front of the body. People can easily notice it is a DJ playing the disc. Other similar dance moves like playing the guitar, bowing at the end, etc. are also incorporated into the dance. Apart from that, we programed the human-like dance to last as long as possible so that the participants observe fewer repeated robot actions while listening to music, because people typically believe that a human would not repeat a dance move in a short time. All of these methods can help the participants in developing an appropriate mental model of the robot. For the robot-like dance, we programmed several simple mechanical moves and repeated them. For example, the Pepper robot can freely move its arms, so we made the action of waving the arm up and down unsymmetrical. Then we made the robot repeat this action several times, and a piece of robot dance action is finished.

- **Voice** Hearing is one of the most important human senses, and a robot's voice can also affect the user's perception of its identity. In this study, the voice was adjusted by the built-in voice settings. The 'naoenu' setting is a distinctly human voice with tone and emotion. The 'naomnc' setting is apparently a synthesized voice which was typically considered to be robot-like (Walters, Syrdal, Koay, et al., 2008).
- **Dialogue** As indicated in Section 2.6, Fischer (2006) has summarised the literature on the distinctions between human-to-human communication (HHC) and human-computer interaction (HCI) in several aspects. Even though the researcher is solely interested in human responses to human or robots, the differences can reflect how people perceive conversations with robot-like robots and subconsciously use the same way of speaking. Inspired by these works, we decided to make the dialogue with a robot-like robot with the following features compared to the human-like robot: less polite, with fewer words per conversation, a smaller lexicon, fewer syntactic, more simplification, include a little over-specification, and technical disclosure rather than emotional disclosure.

## 4.3. Human-robot dialogue

The design of the human-robot dialogue adhered to the instructions in Section 4.2. To properly address the research questions, two additional guidelines were used. Since the duration of the human-robot interaction activity, except for the memory recall part, served as a measure for fluency, it is important that both groups' dialogues have a similar length. Consequently, some basic and general questions were conducted in the beginning, these questions were straightforward, and the responses were predictable. The dialogue was structured to follow this guideline before the memory recall question was raised. As a result, the activity of this segment was used to measure the objective fluency metric.

Another guideline for dialogue design was to make every effort to elicit a story from each participant regarding a song-related memory. During the memory recall task, the participant was encouraged to elaborate on their memory a minimum of once and a maximum of twice.

Based on the guidelines discussed above, the human-robot dialogue was formed (full version in Appendix C). Table 4.1 shows an example segment of the dialogue. It can be noticed that this robot-like robot has a bit of over-specification and considerable technical disclosure.

| Human-like Condition | Robot-like Condition |
|---|---|
| I am an intelligent social robot, and my name is Robin. I am designed to communicate with people and act like a human, just like you! Please make yourself comfortable. Are you ready to begin? | My name is Tronic, and I am a humanoid robot. I am fully equipped to be able to communicate with humankind in a robot way. I am connected to the Internet. I have sensors and much more. Can you hear me well? |

Table 4.1: Example of the human-robot dialogue

## 4.4. Music playlist

The participant would choose a song from a list offered by the robot, and this song would be played in the experiment for both the participant and the robot. In order to do this, a playlist of songs for the experiment was created.

Based on the target group and the research question, there are two guidelines for creating this music playlist. The first guideline is that the music should be highly likely to be familiar to the participants, which can help them to evoke memories associated with it. For this reason, we chose to use a well-known ranking website [5]: Billboard, which increases the possibility that the participant has already heard a song on the list we prepared. The second guideline is that we would like to use the music to evoke autobiographical memories of the participants, so the music should be ”old”. Every ten years, Billboard compiles the most popular songs. Since the expected age of the participants is between 20 and 30 years old, the lists for the 2000s and 2020s are not appropriate for this purpose. Therefore, the ranking list called ”DECADE-END CHARTS Hot 100 Songs in 2010s” in the Billboard was used for preparing the music playlist.

Finally, we reviewed each song on this list, listed their genres based on data from Google, and then categorised them by genre. For this experiment, the four most popular genres were chosen, which were: pop, R&B, electronic and hip hop (see Figure 4.3). To create the music playlist, we chose the top 8 or 9 songs from each of these 4 categories. However, the length of every song varies, to simplify the control variables of the experiment, we trimmed them all to the same length.

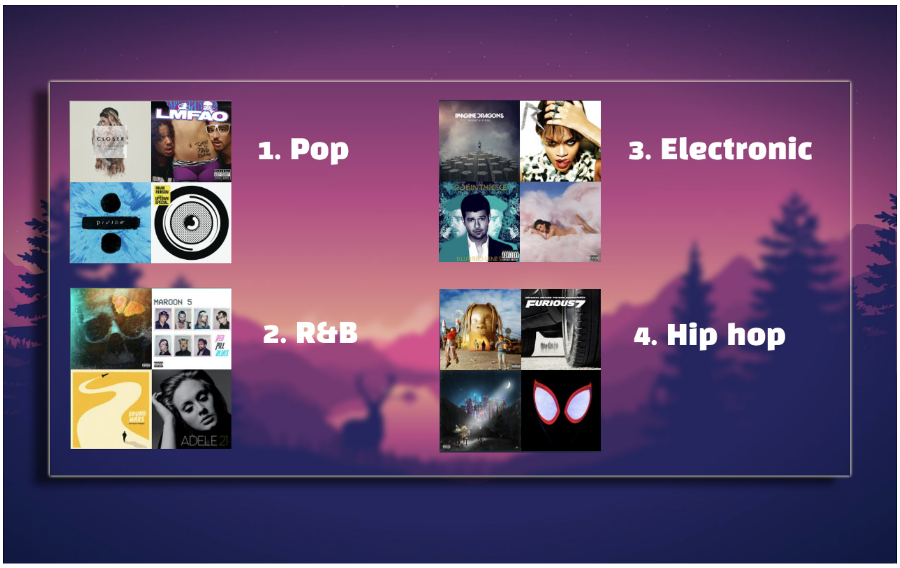



Figure 4.3: Genre selection interface

[5]https://www.billboard.com/

# 5

# Results

The data for this thesis was collected through video observation and a follow-up questionnaire, and the data analysis was conducted in R language. First, the variables for the video observations, the performance of memory recall and the objective fluency metrics, were discussed. The results of the questionnaire's subjective fluency measures and big five personality test were then reviewed. The last part covered a few interesting observations.

## 5.1. Video observations

### 5.1.1. Objective fluency metrics

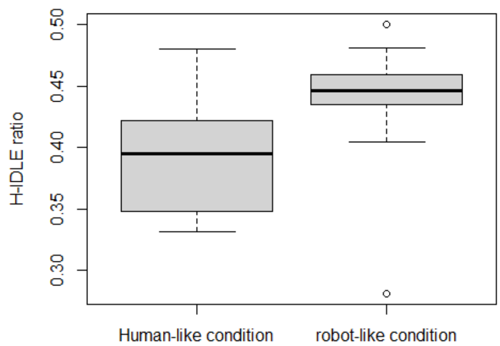



Figure 5.1: H-IDLE ratio from video observations

As was noted in Section 3.6, objective fluency was measured by the H-IDLE time ratio, Figure 5.1 showed the outcomes of this measurement. It is apparent from this figure that the H-IDLE ratio in the human-like condition (Mdn = 0.395, M = 0.393, SD = 0.047) generally lower than in the robot-like condition (Mdn = 0.446, M = 0.438, SD = 0.049).

For these two datasets, we performed a normality test, and the robot-like condition one failed. Thus, the Mann-Whitney test was used to analyse the data, and it revealed that the median H-IDLE ratio for the human-like group differs significantly from that of the robot-like group ($w = 46$, $p < 0.01$).

### 5.1.2. Performance of memory recall

At the end of the experiment, participants were asked to recall memories related to the song they selected, it was determined using the TEMPau test score, which was discussed in Section 3.6. The primary experimenter reviewed all of the video data, and to reduce scoring bias, a second independent rater evaluated 20% of the data (N = 6) as well. The results indicated that they have the same viewpoint regarding the assessment with 100% inter-rater reliability. As shown in Figure 5.2, the TEMPau score of the human-like condition (Mdn = 1, SD = 1.18) is apparently higher than the robot-like condition (Mdn = 1, SD = 0.68). Due to the non-normal distribution of these two datasets, we performed the Mann-Whitney test once more. The results showed that there is a significant difference between the two groups (w = 64, $p < 0.05$).

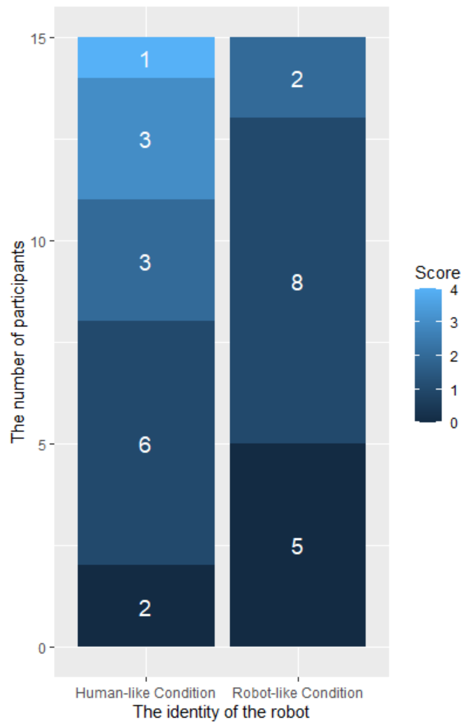



Figure 5.2: TEMPau test score from video observations (range from 0 to 4)

## 5.2. Questionnaire

### 5.2.1. Subjective fluency metrics

The subjective fluency metrics, as described in Section 3.6, consisted of nine questions, each of which employed a 7-point Likert scale. In order to create a more normal distribution, which may be favourable to the t-test, a technique called summed scores was applied (de Winter & Dodou, 2019). Due to the fact that both datasets passed the normality test, we performed a two-sample, two-tail, equal variance t-test. The findings show no statistically significant differences ($t(28) = -0.036$, $p > 0.05$) between the 15 participants who interacted with the human-like robot (M = 48.07, SD = 5.12) and the 15 individuals in the robot-like condition (M = 48.13, SD = 4.85).

### 5.2.2. Big 5 personality

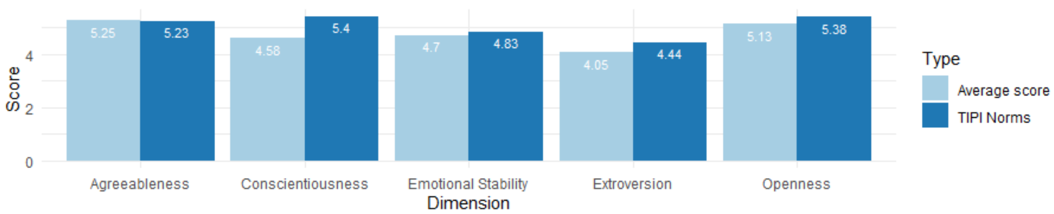



Figure 5.3: Comparison of average scores with TIPI norms

Figure 5.3 shows that the average scores of the 30 participants are generally consistent with the TIPI score Norms. As was discussed in Section 3.7, the Pearson correlation test was used to determine the relationship between the data. The results of the positive correlational analysis are summarised in Figures 5.4A and 5.4B, the TEMPau test scores exhibit a significant but weak positive correlation to emotional stability ($r(28) = 0.398$, $p < 0.05$) and conscientiousness ($r(28) = 0.41$, $p < 0.05$). Participants with higher scores on the personality traits of emotional stability or conscientiousness expressed a richer memory.

Figure 5.4C indicates that the conscientiousness personality trait and the H-IDLE ratio were found to be negatively correlated ($r(28) = -0.34$, $p < 0.05$). In this human-robot interaction task, participants with higher conscientiousness personality trait scores had a lower H-IDLE ratio. The lower H-IDLE ratio means the inactive time ratio of the participant is also lower, which indicates more HRI fluency.

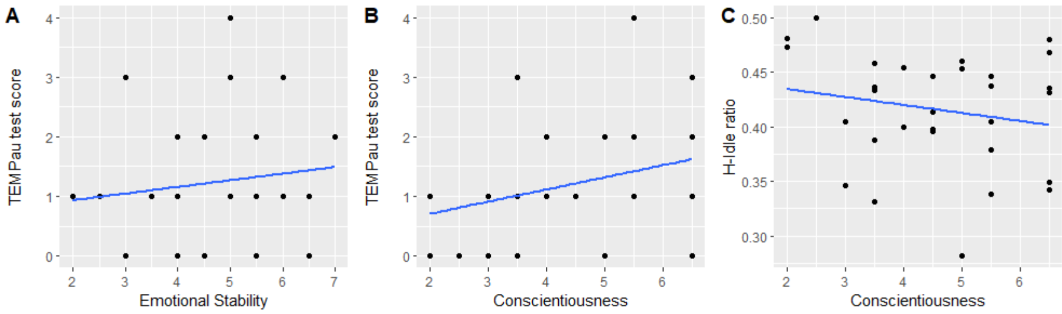



Figure 5.4: The relationship of personality to the H-IDLE ratio and TEMPau test scores (3 significant results)

## 5.3. Other observations

Another intriguing finding from this study was that twice as many individuals in the human-like condition group (8/15) were willing to dance with the robot as those in the robot-like condition (4/15). Even one of them in the human-like group attempted to touch the hand of the robot.

In general, the majority of participants reported having a pleasant experience after the experiment. For example, a participant took a video of the robot and asked if she could have it because she adored the robot. However, one person reported a negative experience, claiming, ”The robot scared me.”

# 6

# Discussion and Conclusion

The constraints of the experiment were discussed first, and then we revisit the research questions. Future work is presented after this section, the study's conclusions are then reached.

## 6.1. Discussion about the experiment

### 6.1.1. Robot platform constraints

The experimental setup is a crucial aspect in deciding the outcome of the experiment. In this part, we will check the design of the experiment to see if there are any components of it that could have a negative or positive impact on the outcomes.

The robot's identity was created by considering numerous factors such as voice, behaviour, appearance, and the dialogue with human. However, some of these factors cannot be fully realised. For example, the robot only has three default voice settings and we select two of these as the human-like condition voice and the robot-like condition voice. The result is that the human-like voice is fine in general, however several participants complained that the voice of the robot condition sounded odd. Some of the behaviour is also constrained by the robot's limited degrees of freedom. For example, under the human-like condition, we intend to employ a tilted head to demonstrate curiosity, unfortunately, the Pepper robot does not enable this behaviour.

### 6.1.2. Running the experiment

Unexpected events that occurred during the experiment's execution may have affected the findings. The participants occasionally attempt to make physical touch with the robot. To avoid hurting people, the Pepper robot contains a safety feature: if someone approaches it too closely, the robot will shut off all of its motors. As a result, the robot may dance strangely, which could affect the participants' perception of the robot.

Another unanticipated problem that occurred during the experiment was that a few participants reported to hear the sound of a mouse clicking behind the curtains, which meant they realised the experimenter was controlling the robot. These sounds may undermine the participants' illusions about the robot's autonomy and thus affect the way they interact with it.

### 6.1.3. Limitations of the experiment

The experiment includes 30 people, which is a fairly small sample size given that some of the data cannot pass the normality test. As mentioned in Chapter 1, we were not able to recruit elderly participants and needed to use students instead. Further experiments with the elderly can be carried out in the future

The Hawthorne effect (Adair, 1984) is a form of reaction in which people alter a behaviour pattern in response to being aware that they are being watched, it might have an impact on the participants' behaviour because this experiment was carried out in a lab. Participants may have behaved differently than usual since they were aware that they were taking part in a study and that their data would be recorded. For instance, when discussing memory recall related to a song, a participant commented, ”This song reminds me of my girlfriend.” But when the robot prompted him to elaborate, he turned to face the camera and declared that it was a secret and that he would prefer not to discuss it.

## 6.2. Revisiting the research questions

The purpose of the experiment is to find answers to the research questions, as presented in Chapter 1. In this section, these research questions are discussed.

### 6.2.1. Which type of robot, robot-like or human-like, produces the most fluent interaction in the elderly care scenario?

This research question was measured by two measures: the objective fluency metrics (H-IDLE ration) and the subjective fluency metrics (questionnaire). Interestingly, the two metrics show different results. The objective metrics reveal that the fluency performance of the human-like condition group is superior to that of the robot-like condition. However the subjective metrics indicate that there is no discernible difference between the two groups. Although participants did not seem to notice objective fluency differences between the two experimental groups, these fluency differences actually existed objectively when individuals interacted with different types of robots. Based on the outcome, the fluency hypothesis from Table 3.2 was rejected.

Kanda et al. (2004)’s study on the relationship between objective outcomes (subjects’ physical motions) and subjective impressions found that these two components may differ in the HRI study. In another major study Laban et al. (2020) discovered that people subjectively feel they disclose more to humans than to human-like robot or disembodied agents, although the results show no real observable differences. These studies demonstrate that objective outcomes can differ from subjective outcomes in HRI. In this study, although the results were also different, they are not the opposite. Thus, we arrived at the conclusion that a robot with human characteristics can communicate with people more fluently overall when engaging in music listening activities.

### 6.2.2. Does the robot-like or human-like robot bring about the best memory performance (MEAMs) in the music listening activity?

The MEAMs performance was measured by the TEMPau test scale. According to the data gathered during the video observations, participants in the human-like condition recall more memories than those in the robot-like condition. We may conclude that the human-like robot performs better on memory recall in the music listening activity. The hypothesis regarding MEAMs is supported by this result.

Additionally, the study of personality discovered a positive link between emotional stability and TEMPau test result, implying that people with better emotional stability want to share more memory with the robot. This result may be explained by the fact that people with higher emotional stability prefer the human-like robot rather than the robot-like robot (Salem et al., 2015; Walters, Syrdal, Dautenhahn, et al., 2008) as indicated in Section 5.4, and in a human-robot interaction exercise, Salem et al. (2015) discovered that participants with higher emotional stability felt closer to the robot. Thus, those with higher emotional stability tend to discuss more about their memory because they feel more connected to the human-like robot, which is consistent with our findings. Previous studies have demonstrated a positive relationship between conscientiousness and the intention to self-disclose (Loiacono et al., 2012). This was also reflected in this study, the participants in the experiment with high conscientiousness personality trait scores had higher TEMPau test scores, which indicates that they disclosed more details during the memory recall phase of the experiment. Thus, the conscientiousness personality trait can be a latent variable for the measurement of the TEMPau test score in the future work.

## 6.3. Future work

Improving the current experiment is essential for future research. Despite the fact that this study was conducted with student volunteers, the target group is the elderly people. As a result, the first step in future study should be to conduct this experiment with the actual target group: the elderly. To accomplish this, some changes must be made. The music playlist should be adjusted to the new target audience, and the robot's voice should also be retested to see if it is appropriate for the elderly.

Furthermore, several enhancements to the experiment should be made in the future work. Some people think the robot-like robot is creepy, thus some alterations may be made to it. For example, the voice and dance of the robot, as reported by the participants. More distinction of the robot identity might also be utilized (e.g., a more vivid voice of the robot), and if possible, a different sort of robot could be used. Pepper is a typical humanoid robot built by SoftBank, it gives participants the idea that it is a human-like robot, which may influence their assessment of the robot's identity.

We proposed that the related future study could focus on the memory recall component. Memory recall or sharing of experiences can be considered as a form of self-disclosure. We only look at one trial of a memory recall experiment in this thesis, and the participants may not trust the robot. However, trust can evolve with time (Vanneste et al., 2014), leading participants to disclose more in the same circumstance as time goes on. The robot can be perceived as non-judgmental, making people more inclined to share their memories with it. As a result, during long-term interaction with the robot, a robot-like robot may perform better in memory recall.

Another intriguing topic that could be explored in future work is the fact that some of the participants danced with the robot during the music listening session. However, the time duration of the dance varies, hence, it can be investigated that in which condition the participants will dance longer, which means the participants enjoy the activity with the robot.

Rapport is also a potential value direction for future work. Rapport, defined by the Oxford dictionary as a friendly relationship in which people understand each other very well, has been acknowledged as a crucial aspect of human interaction. In the field of human-computer interaction, it means "a good feeling of connection and closeness with another" (Zhao et al., 2014). The majority of paper related to this field are focused on human-like behaviour, particularly the virtual agent, which cannot be realized on our robot because it requires small head motions, gaze and facial movement. For example, a complicated model of rapport was developed for their agent in various recent investigations (Papangelis et al., 2014; Zhao et al., 2014). To further the research on the human-robot rapport, Riek et al. (2010) created a realistic head-gesture mimicking robot that can subtly mimic people's head gestures in real time. Furthermore, research (Sinha & Cassell, 2015) on the subject is more likely to be based on long-term investigations. Rapport has not been included in this study because to physical constraints of the system, but it may be in future ones.

Since the experiment in this study was conducted by the Wizard of Oz, a real autonomous robot may be developed in future work, which could enabling long-term interaction in the actual world. However, the autonomous robot necessitates a higher level of equipment's performance, such as large storage space for the user's data, a low latency connection with the computer server or a high-end CPU built-in, a powerful natural language processor to handle the complex real-world conversation, and so on.

## 6.4. Conclusion

In this thesis, we designed, implemented, and evaluated a music assistive robot with two identities. One is attempting to act like a human, which was with intuitive fluid interaction but may provide the user an inappropriate mental model. The other, who had clear expectations but less intuitively smooth interaction, exhibited robot-like traits. According to the findings of this study, the human-like robot generally outperforms the robot-like robot in terms of the fluency in the human-robot interaction and MEAMs performance in the music listening activity. Although the individuals did not perceive the human-like robot's fluency improvement subjectively. Furthermore, we discovered that participants prefer to dance with the human-like robots just as much as they love to discuss their memories with them.

This research contributes to the development of social robots for elderly care, embedded with music that promotes the health and well-being of older adults. The robot can listen to music with the elderly and talk about music-related memories, which can help them have better mental health. The findings of this research may aid in understanding why more and more social robots are being developed to resemble humans since the human-like robot in the experiment performed better. The experiment was carried out on young adults due to time constraints. Therefore, we suggest that future research evaluate the robot using actual elderly people. The results of this study may also serve as a good point for the long-term application of social robots because the music service can be repeatedly used. The elderly can benefit from the music activity by having a wide range of memories stimulated, which can improve their health and well-being. It is valuable and challenging to develop such a social robot which can be used for the long term, interaction and collaboration may be enhanced over time depending on memory acquired from each activity.

# A

# Questionnaire

1. Please fill the tag number
2. How old are you?
3. What do you think this robot behaves like, more like a human or a robot? (1 means fully human-like, 7 means fully robot-like.)
4. The robot is your team member for this music listening event, please rate your team member. (7 Point Likert Scale)
   (a) The human-robot team worked fluently together
   (b) The human-robot team's fluency improved over time.
   (c) The robot contributed to the fluency of the interaction.
   (d) I had to carry the weight to make the human-robot team better.
   (e) The robot contributed equally to the team performance.
   (f) I was the most important team member on the team.
   (g) The robot was the most important team member on the team. I trusted the robot to do the right thing at the right time.
   (h) The robot was trustworthy.
5. Please answer how you see yourself. (7 Point Likert Scale)
   (a) I see myself as extraverted, enthusiastic.
   (b) I see myself as critical, quarrelsome.
   (c) I see myself as dependable, self-disciplined.
   (d) I see myself as anxious, easily upset.
   (e) I see myself as open to new experiences, complex.
   (f) I see myself as reserved, quiet.
   (g) I see myself as sympathetic, warm.
   (h) I see myself as disorganized, careless.
   (i) I see myself as calm, emotionally stable.
   (j) I see myself as conventional, uncreative.

# B

# Consent Form

You are being invited to participate in a research study titled Comparative Performance of different Music Assistive Robots. This study is being done by Yanzhe Li, Mark Neerincx, Frank Broz from the TU Delft.

The purpose of this research study is observed how you interact with the music assistive robot and will take you approximately 30 minutes to complete. The data will be used a master thesis and further publications. We will be asking you to interact with an intelligent robot to listen to music together.

As with any online activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by storing all your data in professionally protected TUDelft storage and pseudo-anonymized. (e.g. blurred videos or pictures) You can withdraw your data for up to 1 month after the data collection, beyond which the data cannot be removed. Your personal research data will be destroyed after the end of the research project.

Your participation in this study is entirely voluntary and you can withdraw at any time by just telling the experimenter. You are free to omit any questions.

PLEASE TICK THE APPROPRIATE BOXES

1. I have read and understood the study information dated 29/07/2022 or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.
3. I understand that taking part in the study involves: A video recording for the experiment and the data will be stored securely on TUDelft storage.
4. I understand that the study will end Aug. 31st 2022
5. I understand that taking part in the study also involves collecting specific personally identifiable information (PII) name, age, and associated personally identifiable research data (PIRD) Video records,
6. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach. The data will be pseudo-anonymized and stored on the Project Storage at TU Delft which has a high standard of security and only the research team can access.
7. I understand that personal information collected about me that can identify me, such as name, the data will be pseudo-anonymized and only the primary researchers will have access to the pseudo-anonymized keys.

8. I understand that after the research study the de-identified information I provide will be used for Master's thesis, further publications.
9. I agree that my responses, views or other input can be quoted anonymously in research outputs
10. I give permission for the de-identified anonymized video records that I provide to be archived in https://webdata.tudelft.nl/staff-umbrella/MasterThesisyanzheli/ repository so it can be used for future research and learning.
11. I understand that access to this repository is restricted only to primary researchers according to the access status to be conferred.

# C

# Full Experimental Protocol

1. Help the participant sign the Consent Form
2. Lead the participant to the seat.
3. Launch the robot and start the dialogue in Table C.1.

| **Human-like Condition** | **Robot-like Condition** |
|---|---|
| • Hi there, would you like to listen to some music with me? | • Hi there, do you want to listen to some music? |
| *If answer No:* OK, that is fine, whenever you want to do it, please let me know. | *If answer No:* OK, that is fine, whenever you want to do it, please let me know. |
| • Great, please allow me to begin by introducing myself. | • Nice, I will start by introducing myself to you. |
| • I am an intelligent social robot, and my name is Robin. I am designed to communicate with people and act like a human, just like you! Please make yourself comfortable. Are you ready to begin? | • My name is Tronic, and I am a humanoid robot. I am fully equipped to be able to communicate with humankind in a robot way. I am connected to the Internet. I have sensors and much more. Can you hear me well? |
| • Great! I bring some candies for you, just on the table, help yourself. I also have a music box on the table, hope you like it too. Recently, I have learned a new ability, I can listen to music with you. Let us have a great time. | • Good! There are candies on the table, help yourself, also a music box here, pretty cool, right? Recently, I have been programmed with a new function, I can play music for you. Let us try it |
| • Before the music starts. May I ask you for some information to create a profile for you? | • Before the music starts. Can I ask you for some information in order to create a profile for you? |
| • May I ask your name? | • May I ask your name? |
| • I like your name. | • That is a nice name! |
| • And, may I know your age? | • And, may I know your age? |
| *If answer No:* OK, that's fine, let's move on to the next question. | *If answer No:* OK, that's fine, let's move on to the next question. |
| • Do you usually listen to music? | • Do you usually listen to music? |
| • What type of music do you like? | • What type of music do you like? |

Table C.1: The human-robot dialogue

| **Human-like Condition** | **Robot-like Condition** |
|---|---|
| • Where do you usually listen to music? | • Where do you usually listen to music? |
| • Have you ever bought an album? | • Have you ever bought an album? |
| *If yes:* Wow, cool. *If no:* Yes, Spotify seems to be used often. | *If yes:* Wow, cool. *If no:* Yes, Spotify seems to be used often. |
| • Hmmm. Ok, great, That is all about creating a profile, I will remember that. | • (blip sounds) Ok, I have created a profile for you. |
| *Pepper display the genre pic on the tablet.* | *Pepper display the genre pic on the tablet.* |
| • Now, it is music time, I selected 4 popular genres for you. | • Now, it is music time, I have 4 popular genres displayed on my tablet, |
| • Please tell me which one do you want? | • Please tell me which one do you want? |
| *The participant selects a genre.* | *The participant selects a genre.* |
| *Back up answer:* Oh, I can not see what you are pointing at with my tablet, you can just tell me. | *Back up answer:* My tablet does not have a touch function, you can just tell me. |
| *The participant selects a song from a list of selected genre.* | *The participant selects a song from a list of selected genre.* |
| *If no clear choice:* Do you want to try the first one? | *If no clear choice:* Do you want to try the first one? |
| *if answer no again* Let me pick one for you. | *if answer no again* Let me pick one for you. |
| • Do you know this song? | • Do you know this song? |
| *If no:* Okay, then, we can choose another one you know, let us start with genre selection. | *If no:* Okay, then, we can choose another one you know, let us start with genre selection. |
| *If yes:* Cool, good choice. Glad to hear you know it. I will find it for you. | *If yes:* Cool, good choice. Glad to hear you know it. I will find it for you. |
| *Music start* | *Music start* |
| • Ok, the music starts now, I can move my body to the music, can you? | • The music starts now. I have the ability to move my body to music, can you? |
| • hum, nice song. | • (bleep sounds), nice song. |
| • I really like this song, do you listen to it often? | • The statistics show that this is a popular song, do you listen to it often? |
| *If Yes:* Who do you usually listen to this song with? | *If Yes:* Who do you usually listen to this song with? |
| *If No:* Who would you listen to this song with? | *If No:* Who would you listen to this song with? |
| • Cool, I like to listen to this with my friends. How does this song make you feel? | • I will record this information in your profile. How does this song make you feel? |
| • What is your favorite part of this song? | • What is your favorite part of this song? |
| • I also really like that part. | • My analysis indicates that is a popular answer. |
| *If no clear answer:* Ok, I understand you. | *If no clear answer:* Ok, I understand you. |
| • What memories does this song bring up for you? | • What memories does this song bring up for you? |

Table C.1: The human-robot dialogue

| Human-like Condition | Robot-like Condition |
|---|---|
| *If no answer:* I understand it does not bring up a memory, but does it remind you of anything or trigger thoughts or feelings? | *If no answer:* I understand it does not bring up a memory, but does it remind you of anything or trigger thoughts or feelings? |
| • Could you add something about the memory? | • Could you add something about the memory? |
| *If not enough:* That's nice to hear, could you tell me a bit more about it? | *If not enough:* Ok, could you tell me a bit more about it? |
| • Well, that is all, thank you for participating in our experiment. | • (blip sounds) That is all, thank you for participating in our experiment. |

Table C.1: The human-robot dialogue

4. Serve a survey after the conversation is over.

# Bibliography


Aaltonen, I., Arvola, A., Heikkilä, P., & Lammi, H. (2017). Hello pepper, may i tickle you? children's and adults' responses to an entertainment robot at a shopping mall. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 53–54. https://doi.org/10.1145/3029798.3038362

Abe, N., Maguire-Rosier, K., & Andreallo, F. (2020, July 12). *Pilot classification of human-robot dance performances*. https://doi.org/10.13140/RG.2.2.16853.86243

Adair, J. G. (1984). The hawthorne effect: A reconsideration of the methodological artifact [Place: US Publisher: American Psychological Association]. *Journal of Applied Psychology*, *69*(2), 334–345. https://doi.org/10.1037/0021-9010.69.2.334

Allegra, D., Alessandro, F., Santoro, C., & Stanco, F. (2018). Experiences in using the pepper robotic platform for museum assistance applications [ISSN: 2381-8549]. *2018 25th IEEE International Conference on Image Processing (ICIP)*, 1033–1037. https://doi.org/10.1109/ICIP.2018.8451777

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, *38*(4), 547–566. https://doi.org/10.1006/imms.1993.1026

Baird, A., & Thompson, W. F. (2018). The impact of music on the self in dementia [Publisher: IOS Press]. *Journal of Alzheimer's Disease*, *61*(3), 827–841. https://doi.org/10.3233/JAD-170737

Bandini, S., Gasparini, F., & Giltri, M. (2019). Personalized music experience for the wellbeing of elderly people, 335–340. https://doi.org/10.1007/978-3-030-34770-3_28

Baraglia, J., Cakmak, M., Nagai, Y., Rao, R., & Asada, M. (2016). Initiative in robot assistance during collaborative task execution [ISSN: 2167-2148]. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 67–74. https://doi.org/10.1109/HRI.2016.7451735

Belfi, A. M., Karlan, B., & Tranel, D. (2016). Music evokes vivid autobiographical memories [Publisher: Routledge _eprint: https://doi.org/10.1080/09658211.2015.1061012]. *Memory*, *24*(7), 979–989. https://doi.org/10.1080/09658211.2015.1061012

Cakmak, M., Srinivasa, S. S., Lee, M. K., Kiesler, S., & Forlizzi, J. (2011). Using spatial and temporal contrast for fluent robot-human hand-overs. *Proceedings of the 6th international conference on Human-robot interaction*, 489–496. https://doi.org/10.1145/1957656.1957823

Chao, C., & Thomaz, A. L. (2012). Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets [Accepted: 2012-08-31T19:39:58Z Publisher: Georgia Institute of Technology]. https://doi.org/10.5898/JHRI.1.1.Chao

Cruz-Maya, A., & Tapus, A. (2016). Influence of user's personality on task execution when reminded by a robot. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social robotics* (pp. 829–838). Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_81

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies — why and how. *Knowledge-Based Systems*, *6*(4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-N

Damholdt, M. F., Nørskov, M., Yamazaki, R., Hakli, R., Hansen, C. V., Vestergaard, C., & Seibt, J. (2015). Attitudinal change in elderly citizens toward social robots: The role of personality traits and beliefs about robot functionality. *Frontiers in Psychology*, *6*. Retrieved August 19, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01701

Darling, K. (2021, April 20). *The new breed: What our history with animals reveals about our future with robots* [Google-Books-ID: oTrGDwAAQBAJ]. Henry Holt; Company.

de Winter, J. F., & Dodou, D. (2019). Five-point likert items: T test versus mann-whitney-wilcoxon (*Addendum added October 2012*). *Practical Assessment, Research, and Evaluation*, *15*(1). https://doi.org/https://doi.org/10.7275/bj1p-ts64

Douka, S., Zilidou, V. I., Lilou, O., & Manou, V. (2019). Traditional dance improves the physical fitness and well-being of the elderly [Publisher: Frontiers Media S.A]. *Frontiers in Aging Neuroscience*, *11*, 75. https://doi.org/10.3389/FNAGI.2019.00075/BIBTEX

Dragan Anca D., Bauman Shira, Forlizzi Jodi, & Srinivasa Siddhartha S. (2015). Effects of robot motion on human-robot collaboration, 51–58.

Duffy, B. R. (2003). Anthropomorphism and the social robot [Publisher: North-Holland]. *Robotics and Autonomous Systems*, *42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

El Haj, M., Fasotti, L., & Allain, P. (2012). The involuntary nature of music-evoked autobiographical memories in alzheimer's disease. *Consciousness and Cognition*, *21*(1), 238–246. https://doi.org/10.1016/j.concog.2011.12.005

El Haj, M., Gandolphe, M. C., Gallouj, K., Kapogiannis, D., & Antoine, P. (2018). From nose to memory: The involuntary nature of odor-evoked autobiographical memories in alzheimer's disease. *Chemical Senses*, *43*(1), 27–34. https://doi.org/10.1093/chemse/bjx064

El Haj, M., Kapogiannis, D., & Antoine, P. (2016). Phenomenological reliving and visual imagery during autobiographical recall in alzheimer's disease. *Journal of Alzheimer's disease : JAD*, *52*(2), 421–431. https://doi.org/10.3233/JAD-151122

El Haj, M., Postal, V., & Allain, P. (2012). Music enhances autobiographical memory in mild alzheimer's disease [Publisher: Routledge _eprint: https://doi.org/10.1080/03601277.2010.515897]. *Educational Gerontology*, *38*(1), 30–41. https://doi.org/10.1080/03601277.2010.515897

Erebak, S., & Turgut, T. (2019). Caregivers' attitudes toward potential robot coworkers in elder care. *Cognition, Technology & Work*, *21*(2), 327–336. https://doi.org/10.1007/s10111-018-0512-0

Fischer, K. (2006). What computer talk is and isn't [Publisher: AQ]. *Human-Computer Conversation as Intercultural Communication*, *17*.

Fischer, K. (2019). Why collaborative robots must be social (and even emotional) actors. *Techné: Research in Philosophy and Technology*, *23*(3), 270–289. https://doi.org/10.5840/techne20191120104

Frijns, H. A., & Schürer, O. (2020). The impact of human-like and device-like signals on people's mental models of robots.

Fu, M. C., Belza, B., Nguyen, H., Logsdon, R., & Demorest, S. (2018). Impact of group-singing on older adult health in senior living communities: A pilot study [Publisher: Elsevier]. *Archives of Gerontology and Geriatrics*, *76*, 138–146. https://doi.org/10.1016/J.ARCHGER.2018.02.012

Gockley, R., & MatariĆ, M. J. (2006). Encouraging physical therapy compliance with a hands-off mobile robot. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 150–155. https://doi.org/10.1145/1121241.1121268

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation, 55–60. https://doi.org/10.1109/ROMAN.2003.1251796

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Goyal, S., Maruping, L., & Robert, L. (2008). Diversity and conflict in teams: A faultline model perspective. [Publisher: Academy of Management]. *Academy of Management Proceedings*, *2008*(1), 1–6. https://doi.org/10.5465/ambpp.2008.33640695

Graaf, M., Ben Allouch, S., & van Dijk, J. (2017). Why do they refuse to use my robot? In B. Mutlu, M. Tscheligi, A. Weiss, & J. E. Young (Eds.). ACM. https://doi.org/10.1145/2909824.3020236

Gupta, A., Bridges, N., & Kamino, W. (2021). Musically assistive robot for the elderly in isolation [Place: New York, NY, USA Publisher: ACM]. *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 620–621. https://doi.org/10.1145/3434074

Gustafson, J. (2002). *Developing multimodal spoken dialogue systems: Empirical studies of spoken human–computer interaction* (Doctoral dissertation). KTH.

Hall, C. S., & Lindzey, G. (1957). *Theories of personality* [Pages: xi, 572]. John Wiley & Sons Inc. https://doi.org/10.1037/10910-000

Hannah Richardson. (2017). Robots could help solve social care crisis, say academics. *BBC News*. Retrieved November 13, 2021, from https://www.bbc.com/news/education-38770516

Haring, K. S., Watanabe, K., Velonaki, M., Tossell, C. C., & Finomore, V. (2018). FFAB—the form function attribution bias in human–robot interaction. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 843–851. https://doi.org/10.1109/TCDS.2018.2851569

Hays, T., & Minichiello, V. (2005). The meaning of music in the lives of older people: A qualitative study [Publisher: SAGE Publications Ltd]. *Psychology of Music*, *33*(4), 437–451. https://doi.org/10.1177/0305735605056160

Hitzenberger, L., & Womser-Hacker, C. (1995). Experimentelle untersuchungen zu multimodalen natürlichsprachlichen dialogen in der mensch-computer-interaktion. sprache und datenverarbeitung, *Sprache und Datenverarbeitung*, *19*, 51–61.

Hoffman, G. (2019). Evaluating fluency in human–robot collaboration [Publisher: IEEE]. *IEEE Transactions on Human-Machine Systems*, *49*(3), 209–218.

Hoffman, G., & Breazeal, C. (2007). Cost-based anticipatory action selection for human–robot fluency. *IEEE Transactions on Robotics*, *23*(5), 952–961. https://doi.org/10.1109/TRO.2007.907483

Huang, C.-M., Cakmak, M., & Mutlu, B. (2015). Adaptive coordination strategies for human-robot handovers. https://doi.org/10.15607/RSS.2015.XI.031

Isaacson, S., Rice, G., & Boerkoel Jr, J. C. (2020). MAD-TN: A tool for measuring fluency in human-robot collaboration. *arXiv:1909.06675 [cs]*. Retrieved November 24, 2021, from http://arxiv.org/abs/1909.06675

Jakubowski, K., & Ghosh, A. (2021). Music-evoked autobiographical memories in everyday life [Publisher: SAGE Publications Ltd]. *Psychology of Music*, *49*(3), 649–666. https://doi.org/10.1177/0305735619888803

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories [Publisher: Routledge _eprint: https://doi.org/10.1080/09658210701734593]. *Memory*, *15*(8), 845–860. https://doi.org/10.1080/09658210701734593

Jäncke, L. (2008). Music, memory and emotion. *Journal of Biology*, *7*(6), 21. https://doi.org/10.1186/jbiol82

Kanda, T., Ishiguro, H., Imai, M., & Ono, T. (2004). Development and evaluation of interactive humanoid robots [Conference Name: Proceedings of the IEEE]. *Proceedings of the IEEE*, *92*(11), 1839–1850. https://doi.org/10.1109/JPROC.2004.835359

Kiesler, S. (2005). Fostering common ground in human-robot interaction, 729–734. https://doi.org/10.1109/ROMAN.2005.1513866

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. https://doi.org/10.1145/506443.506491

Kok, R., Rothweiler, J., Scholten, L., van Zoest, M., Boumans, R., & Neerincx, M. (2018). Combining social robotics and music as a non-medical treatment for people with dementia [Publisher: Institute of Electrical and Electronics Engineers Inc]. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 465–467. https://doi.org/10.1109/ROMAN.2018.8525813

Kompatsiari, K., Tikhanoff, V., Ciardo, F., Metta, G., & Wykowska, A. (2017). The importance of mutual gaze in human-robot interaction. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (Eds.), *Social robotics* (pp. 443–452). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_44

Krause, J. (1992). *Computer talk* [Google-Books-ID: y3cbAAAAIAAJ]. G. Olms.

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots [ISSN: 2167-2148]. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. https://doi.org/10.1109/HRI.2016.7451807

Laban, G., Morrison, V., & Cross, E. S. (2020). Let's talk about it! subjective and objective disclosures to social robots. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 328–330. https://doi.org/10.1145/3371382.3378252

Lee, S.-l., Lau, I. Y.-m., Kiesler, S., & Chiu, C.-Y. (2005). Human mental models of humanoid robots [ISSN: 1050-4729]. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2767–2772. https://doi.org/10.1109/ROBOT.2005.1570532

Li, N., Barrick, M. R., Zimmerman, R. D., & Chiaburu, D. S. (2014). Retaining the productive employee: The role of personality [Publisher: Academy of Management]. *Academy of Management Annals*, *8*(1), 347–395. https://doi.org/10.5465/19416520.2014.890368

Ligthart, M., Henkemans, O. B., Hindriks, K., & Neerincx, M. A. (2017). Expectation management in child-robot interaction [ISSN: 1944-9437]. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 916–921. https://doi.org/10.1109/ROMAN.2017.8172412

Loiacono, E., Carey, D., Misch, A., Spencer, A., & Speranza, R. (2012). Personality impacts on self-disclosure behavior on social networking sites. *AMCIS 2012 Proceedings*. https://aisel.aisnet.org/amcis2012/proceedings/HCIStudies/6

Lou, M.-F., & Mn, R. N. (2001). The use of music to decrease agitated behaviour of the demented elderly: The state of the science [Publisher: John Wiley & Sons, Ltd]. *Scandinavian Journal of Caring Sciences*, *15*(2), 165–173. https://doi.org/10.1046/J.1471-6712.2001.00021.X

Maniadakis, M., Hourdakis, E., & Trahanias, P. (2017). Time-informed task planning in multi-agent collaboration [Publisher: Elsevier]. *Cognitive Systems Research*, *43*, 291–300. https://doi.org/10.1016/j.cogsys.2016.09.004

Mileski, M., Brooks, M., Kirsch, A., Lee, F., Levieux, A., & Ruiz, A. (2019). Positive physical and mental outcomes for residents in nursing facilities using music: A systematic review [Publisher: Dove Press]. *Clinical Interventions in Aging*, *14*, 301. https://doi.org/10.2147/CIA.S189486

Morel, M.-A. (1989). Computer-human communication [Publisher: North-Holland Elsevier Amsterdam]. *The Structure of Multimodal Communication*, 323–330.

Moreno-Morales, C., Calero, R., Moreno-Morales, P., & Pintado, C. (2020). Music therapy in the treatment of dementia: A systematic review and meta-analysis [Publisher: Frontiers Media S.A]. *Frontiers in Medicine*, *7*, 1–11. https://doi.org/10.3389/FMED.2020.00160/BIBTEX

Munawar, K., Kuhn, S. K., & Haque, S. (2018). Understanding the reminiscence bump: A systematic review [Publisher: Public Library of Science]. *PLOS ONE*, *13*(12), e0208595. https://doi.org/10.1371/journal.pone.0208595

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nikolaidis, S., & Shah, J. (n.d.). Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy, 33–40. https://doi.org/10.1109/HRI.2013.6483499

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversational interfaces: Modeling speech convergence with animated personas [Place: New York, NY, USA Publisher: Association for Computing Machinery]. *ACM Trans. Comput.-Hum. Interact.*, *11*(3), 300–328. https://doi.org/10.1145/1017494.1017498

Paepcke, S., & Takayama, L. (2010). Judging a bot by its cover: An experiment on expectation setting for personal robots [ISSN: 2167-2148]. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 45–52. https://doi.org/10.1109/HRI.2010.5453268

Papangelis, A., Zhao, R., & Cassell, J. (2014). Towards a computational architecture of dyadic rapport management for virtual agents. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Intelligent virtual agents* (pp. 320–324). Springer International Publishing. https://doi.org/10.1007/978-3-319-09767-1_41

Peeters, M. M. M., Harbers, M., & Neerincx, M. A. (2016). Designing a personal music assistant that enhances the social, cognitive, and affective experiences of people with dementia [Publisher: Pergamon]. *Computers in Human Behavior*, *63*, 727–737. https://doi.org/10.1016/J.CHB.2016.06.003

Pfadenhauer, M., & Dukat, C. (2015). Robot caregiver or robot-supported caregiving? [Publisher: Springer]. *International Journal of Social Robotics*, *7*(3), 393–406. https://doi.org/10.1007/s12369-015-0284-0

Piolino, P., Desgranges, B., Clarys, D., Guillery-Girard, B., Taconnat, L., Isingrini, M., & Eustache, F. (2006). Autobiographical memory, autonoetic consciousness, and self-perspective in aging [Place: US Publisher: American Psychological Association]. *Psychology and Aging*, *21*(3), 510–525. https://doi.org/10.1037/0882-7974.21.3.510

Platz, F., Kopiez, R., Hasselhorn, J., & Wolf, A. (2015). The impact of song-specific age and affective qualities of popular songs on music-evoked autobiographical memories (MEAMs) [Publisher: SAGE Publications Ltd]. *Musicae Scientiae*, *19*(4), 327–349. https://doi.org/10.1177/1029864915597567

Porzel, R., & Baudis, M. (2004). The tao of chi: Towards effective human-computer interaction. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, 209–216.

Pot, E., Monceaux, J., Gelin, R., & Maisonnier, B. (2009). Choregraphe: A graphical tool for humanoid robot programming [ISSN: 1944-9437]. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 46–51. https://doi.org/10.1109/ROMAN.2009.5326209

Powers, A., & Kiesler, S. (2006). The advisor robot. In M. A. Goodrich, A. C. Schultz, & D. J. Bruemmer (Eds.). ACM Press. https://doi.org/10.1145/1121241.1121280

Preeti Ramaraj. (2021). Robots that help humans build better mental models of robots. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(18), 15730–15731. https://ojs.aaai.org/index.php/AAAI/article/view/17862

Rahman, S. M. M. (2018). Cyber-physical-social system between a humanoid robot and a virtual human through a shared platform for adaptive agent ecology. *IEEE/CAA Journal of Automatica Sinica*, *5*(1), 190–203. https://doi.org/10.1109/JAS.2017.7510760

Richards, M., & Underwood, K. (1984). How should people and computers speak to each other. *Proceedings of Interact*, *84*, 33–36.

Riek, L. D., Paul, P. C., & Robinson, P. (2010). When my robot smiles at me:enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, *3*(1), 99–108. https://doi.org/10.1007/s12193-009-0028-2

Rietz, F., Sutherland, A., Bensch, S., Wermter, S., & Hellström, T. (2021). WoZ4u: An open-source wizard-of-oz interface for easy, efficient and robust HRI experiments. *Frontiers in Robotics and AI*, *8*, 213. https://doi.org/10.3389/frobt.2021.668057

Rothstein, N., Kounios, J., Ayaz, H., & Visser, E. J. (2020). Assessment of human-likeness and anthropomorphism of robots: A literature review. In H. Ayaz & U. Asgher (Eds.), *Advances in neuroergonomics and cognitive engineering. proceedings of the AHFE 2020 virtual conferences on neuroergonomics and cognitive engineering, and industrial cognitive ergonomics and engineering psychology, july 16-20, 2020, USA / hasan ayaz, umer asgher, editors* (pp. 190–196). Springer. https://doi.org/10.1007/978-3-030-51041-1_26

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 141–148. https://doi.org/10.1145/2696454.2696497

Salem, M., Rohlfing, K., Kopp, S., & Joublin, F. (2011). A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction [ISSN: 1944-9437]. *2011 RO-MAN*, 247–252. https://doi.org/10.1109/ROMAN.2011.6005285

Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics*, *8*(2), 303–317. https://doi.org/10.1007/s12369-015-0323-x

Schramm, L. T., Dufault, D., & Young, J. E. (2020). Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design. In T. Belpaeme, J. Young, H. Gunes, & L. Riek (Eds.). ACM. https://doi.org/10.1145/3371382.3378280

Schulkind, M. D., Hennis, L. K., & Rubin, D. C. (1999). Music, emotion, and autobiographical memory: They're playing your song. *Memory & Cognition*, *27*(6), 948–955. https://doi.org/10.3758/bf03201225

Sharma, A., Rathi, Y., Patni, V., & Sinha, D. K. (2021). A systematic review of assistance robots for elderly care, 1–6. https://doi.org/10.1109/ICCICT50803.2021.9510142

Sinha, T., & Cassell, J. (2015). We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, 13–20. https://doi.org/10.1145/2823513.2823516

Skingley, A., & Bungay, H. (2013). The silver song club project: Singing to promote the health of older people [Publisher: MA Healthcare London]. *http://dx.doi.org/10.12968/bjcn.2010.15.3.46902*, *15*(3), 135–140. https://doi.org/10.12968/BJCN.2010.15.3.46902

Softbank. (2022). *Pepper the humanoid and programmable robot | SoftBank robotics*. Retrieved April 15, 2022, from https://www.softbankrobotics.com/emea/en/pepper

Unhelkar, V. V., Siu, H. C., & Shah, J. A. (2014). Comparative performance of human and mobile robotic assistants in collaborative fetch-and-deliver tasks. https://doi.org/10.1145/2559636.2559655

United Nations Department of Economics and Social Affairs Population Division. (2020). World population ageing 2020 highlights. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd-2020_world_population_ageing_highlights.pdf

Vanneste, B. S., Puranam, P., & Kretschmer, T. (2014). Trust over time in exchange relationships: Meta-analysis and theory [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.2198]. *Strategic Management Journal*, *35*(12), 1891–1902. https://doi.org/10.1002/smj.2198

Walters, M. L., Syrdal, D. S., Koay, K. L., Dautenhahn, K., & te Boekhorst, R. (2008). Human approach distances to a mechanical-looking robot with different robot voice styles [ISSN: 1944-9437]. *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 707–712. https://doi.org/10.1109/ROMAN.2008.4600750

Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, *24*(2), 159–178. https://doi.org/10.1007/s10514-007-9058-3

Ward, K., & Heeman, P. A. (2000). Acknowledgments in human-computer interaction. *1st Meeting of the North American Chapter of the Association for Computational Linguistics*.

Wattanasoei, S., Binson, B., Kumar, R., & Somrongthong, R. (2017). Quality of life through listening music among elderly people in semi-urban area, thailand. *J Ayub Med Coll Abbottabad*, *29*(1), 21–26.

WHO. (2021). *Ageing and health*. Retrieved December 11, 2021, from https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2015). *Engineering psychology and human performance* (Fifth edition / Christopher D. Wickens, William S. Helton, Justin G. Hollands, Simon Banbury). Routledge. https://doi.org/10.4324/9781003177616

Zhao, R., Papangelis, A., & Cassell, J. (2014). Towards a dyadic computational model of rapport management for human-virtual agent interaction. In T. Bickmore, S. Marsella, & C. Sidner (Eds.), *Intelligent virtual agents* (pp. 514–527). Springer International Publishing. https://doi.org/10.1007/978-3-319-09767-1_62