Grant agreement No: 101017008

Harmony

Assistive robots for healthcare

Enhancing Healthcare with Assistive Robotic Mobile Manipulation

(HARMONY) | H2020-ICT-2018-20| RIA

Start of the project: 01.01.2021
Duration: 42 months

Deliverable Number	8.3
Deliverable Name	Validated set of multimodal social robot intent behaviours and whole-body behaviours
WP Number	8
Lead Beneficiary	UT
Dissemination Level	Public
Internal Reviewer	IDM
Due Date	30-06-2023
Date of Submission	04-07-2023
Version	2.0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017008

Revision History

Version	Date	Author(s)	Comments
0.1	26-05-2023	Bob Schadenberg	Outline
0.2	28-06-2023	Bob Schadenberg, Hideki Garcia Goo, Jan Kolkmeier, Luisa Fernandes, Theresa Höfker	First draft
1.0	03-07-2023	Bob Schadenberg, Hideki Garcia Goo, Jan Kolkmeier, Luisa Fernandes, Theresa Höfker	Final version
2.0	22-12-2023	Bob Schadenberg, Hideki Garcia Goo, Jan Kolkmeier, Luisa Fernandes, Theresa Höfker, Laura Ermers	Updated version: new sections (4, 6 and 7), and updated section 5.

Table of Contents

1. Summary

In this deliverable, we delineate three research studies on robot behaviour design that we conducted. Additionally, we provide a description of a study slated to run in January and February 2024, outlining the behaviours crafted to facilitate the IDM Harmony robot in executing its task of delivering test samples within the hospital and the description of the development of a novel way of communication, shapeshifting. The overall goal of this deliverable is to describe the efforts taken into the design of the IDM Harmony robot's communicative behaviours. The goal of the first study we discuss is to investigate how well musical utterances (brief pieces of music) can communicate intent and evoke empathy and prosocial behaviour in people. UT designed these musical utterances for the IDM Harmony robot together with two sound designers. The sounds were evaluated in an online study in three different scenarios. While the musical utterances made it significantly easier for participants to interpret the robot's "experienced" emotions (cognitive empathy), we found no evidence that musical utterances improved participants' interpretation of the robot's behaviour (legibility) nor that they were more likely to help it. The latter can also be explained by a ceiling effect on legibility, where the scenario was insufficiently ambiguous. Given our experience and related works, we conclude that it is difficult to communicate intent through musical utterances beyond communicating emotions.

In version 2.0, we included a study where we investigated the effect of age on people's ability to categorise emotions in SFUs. A pre-study was conducted to identify audio stimuli to use in the main study and to assess which type of utterances participants preferred for a hospital robot; listeners highlighted the similarity to natural language, the need for a distinction between human and robot, and their expectations of what a hospital robot should sound like as factors that influence their choice. In the main study, participants were adults of ages 18 to 89, who interpreted emotions expressed through two types of SFU with varying humanlikeness, during an online emotion categorization task. We also considered how other factors influenced listeners' emotion recognition, such as type of emotion, SFU, listeners' gender, and their experience with robots. Results confirm an effect of a decrease of emotion recognition performance on SFUs as age increases, as well as effects on emotion category and SFU database. Additionally, people exhibit preferences for particular types of utterances concerning the robot's appearance, role, and environment. Consequently, further investigation is required to develop a distinct "voice" for IDM's Harmony robot that aligns with people's expectations of its sound and ensures inclusivity for older adults to understand it.

In the third study, we looked for inspiration for robot behaviours together with performers aimed at mitigating the negative effects when the robot makes a mistake. In this last version (2.0) we updated this section with the results of the main study. In this deliverable, we report on how we approach the design of these behaviours together with the

performers and show preliminary results of a pilot and the main study. The performers are enacting scenarios in which they are the robot that needs to resolve a certain social interaction. We will analyse the behaviour of the performers and abstract away from their literal movement to more universal descriptions thereof. These behaviours will then be used as inspiration for the creation of the IDM Harmony robot's behaviour and robot error scenarios.

Then, we report on the behaviours that UT and IDM developed to enable the upcoming user studies with the IDM Harmony robot. These studies will focus on the interactions that can take place with the various stakeholders inside the hospital hallways when the robot is delivering test samples.

Lastly, in this last update (2.0), we report the design of an online study that aims to inform the design of the IDM Harmony robot's recovery strategies when it makes a social mistake. During the study, participants will watch video stimuli depicting different social errors and recovery strategies made by the Harmony robot while it navigates a hospital corridor. The videos will be created in a virtual environment. Participants will rate the severity of the mistake

2. Introduction

The use-cases of the Harmony robot require it to operate in the hospital laboratories and hallways. These are spaces occupied by people who are working, visitors, or patients who come to the hospital for some kind of treatment. As such, the robot will need to carry out its tasks in a manner that is acceptable to those people. At times the robot may also need to proactively engage with people near it to get them to do something it needs so that it can proceed with its primary task; delivering or handling goods within the hospital. For the robot's communication, it then becomes important that people understand what it is trying to communicate (legibility), that it communicates pleasantly, and that the communication is effective in achieving its goal (e.g., getting help from people, or ending an interaction so that it can go on its way to deliver goods). In this deliverable, we report on two studies that relate to the behaviour design for the IDM Harmony robot.

The first study (Section 3), that we report on, relates to the voice design of the robot. In this study we specifically look into the use of musical utterances (brief pieces of music) as a way of communication. Compared to using semantic speech, this type of verbal communication could be particularly useful because it (a) does not rely on understanding language, (b) does not signal that the robot can also understand language itself, and (c) may bias people less because there is no gender associated with the sounds. Compared to other types of sounds, musical utterances may also be more pleasant to hear than all kinds of "beeps and boops" (which can also be used for communication) that are made by various machines inside the hospital. Previous work also showed that musical utterances can be used

to communicate emotions well, which may be useful given the relation between emotions, empathy, and prosocial behaviour: evoking emotions can lead to empathy, which in turn can lead to people engaging in prosocial behaviour (i.e., helping the robot when it needs help).

Additionally, we also ran a second study (Section 4) to investigate the effect of age on the emotion interpretation of different SFU databases containing two different types of SFU and to study the match between them and the IDM Harmony robot. Based on the results of a pre-study which showed that participants preferred the robot to use more “robot-like” gibberish speech, we also manipulated selected audio from the gibberish speech database to make it sound more robot-like. Results showed a decrease in emotion categorization ability across the three databases, and a different preference of the robot speech type across the robot’s appearance, role, and environment.

For the third study (Section 5), we focus on developing robot behaviour that can be used to mitigate the negative effects of a robot making a social mistake (e.g., invading someone’s personal space) by using the IDM Harmony robot’s available modalities and taking into account its limitations (e.g., the absence of verbal communication). These are bound to happen, and it then becomes important to recover from them so that people don't lose trust in the robot, or otherwise feel negatively about it.

For both studies, we incorporated the expertise of artists in the design of the behaviour. For the first study, the musical utterances were designed by sound designers. And for the second study, the multi-modal behaviours came out of sessions with contemporary and hip-hop dancers, and improvisation actors. These artists are experts at translating intent into behaviour. Through these studies, we also aim to learn more about how we can tap into their expertise and creativity to get to compelling and effective robot behaviours.

Then, we report on the robot behaviours that we developed to enable the robot to resolve social situations that can occur while it is navigating the hospital hallways. These can include encountering an obstruction, either because people are blocking its path or an object is in the way. To resolve this event, the robot can interact with the people blocking it by signalling that it needs to pass. Or in case the path is blocked by an object, the robot could ask for help and signal that it needs to pass the object. In addition to behaviours to resolve certain social events, we also developed some other behaviours to improve functionality and safety during user studies. We describe all the behaviours UT and IDM designed and the events in Section 6.

Section 7 describes a study where we create and evaluate different social errors and recovery strategies with the use of different modalities design (minimal, complex, and with the use of sound). This study is scheduled to take place during winter 2024 and its results will be shown in Deliverable 8.5.

3. Musical utterances to communicate intent and evoke empathy in people

3.1 Introduction

In the hospital hallways, situations might occur in which the robot requires some form of support or help from the people in its environment. For example when a door that the robot needs to go through is closed. To resolve such situations, the robot needs to interact with people in such a way that they are on the one hand, willing to help and cooperate with the robot (i.e., showing *prosocial behaviour*) and on the other hand also understand what the robot needs (i.e., the robot's communication is *legible*).

In the past years, semantic free utterances (SFU) gained a lot of attention to be implemented in robots as a means of communication. In short, SFUs can be described as sounds that are made by machines that do not make use of any semantic meaning but still communicate intentions and emotions (Yilmazyildi et al., 2016). Movies like WALL-E and Star Wars with the robot R2D2 demonstrate well that these robots can communicate with and to humans by using sounds and not actual words. One of the benefits of SFUs is that they do not rely on the language understanding of the person interacting with it. And, within a hospital context, we can assume that not everyone will speak the same language. Moreover, communicating through SFUs does not elicit the expectation that the robot can understand the language itself. In that sense, it helps align the expectations of people with the capabilities of the Harmony robot. Lastly, compared to using a human voice, SFUs might be less prone to eliciting (negative) gender stereotypes when interacting with the robot.

A specific type of SFUs are musical utterances. These are musical pieces that result from the variation of different attributes of music, like rhythm, dynamics, timbre, or pitch (Yilmazyildi et al., 2016). This particular type of SFUs could be particularly useful for evoking prosocial behaviour in people through eliciting *empathy*, given that music excels at communicating emotions to the listeners (e.g., Jee et al., 2007). However, communicating anything beyond emotions can be difficult with musical utterances (Savory, Rose & Weinberg, 2019). This is a more general challenge of SFUs, where the sounds in isolation are difficult to interpret (Read & Belpaeme, 2014). However, as previous work from our department at the UT shows, the situational context can give meaning to them (Zaga et al., 2017).

While certain types of SFUs have been shown to convey intent in certain situations, musical utterances have only been regarded as a means to communicate emotions. Moreover, it is not clear whether a successful communication of the robot's emotions also triggers empathy in a person which then results in prosocial behaviour when the robot requires help. In our

study, we will therefore investigate to what extent musical utterances can communicate intentions and emotions. As a secondary objective, we also measured the cognitive and affective empathy and prosocial behaviour evoked by the videos. These measures were included to gain insight into these two constructs for a potential follow-up study. However, the affective empathy and prosocial behaviour measures should be interpreted carefully given that their validity is limited due to the method that we employed. To design the musical utterances, we collaborated with two sound designers. The robot would then use these sounds in three scenarios that may occur in a hospital hallway, which can be resolved by a person assisting the robot. We recorded such instances where the robot would use a musical utterance to communicate with people near it and assessed these videos in an online study using the Prolific recruitment platform.

3.2 Sound and Video Design

3.2.1 Sound Designers

Two sound designers were recruited through convenience. These sound designers were both male and had seven years of experience with electronic music production. Their main experience lies within the digital synthesis of sound, meaning the creation of sound by a computer. The sound designers mainly produce music in the genres of dance music, house, techno, ambient, and funk.

3.2.2 Sound Design Session

For the sound design session, we created a scenario outline. The selection of the scenarios was based on possible situations the robot could encounter while navigating the hospital hallways. These included people blocking the robot's way, a door being closed in the way of the robot, the robot getting stuck on a cable on the floor, and the robot being unable to localise itself. The scenarios were written as design fiction. This consists of something that creates a story world and prototypes reflecting the story world that together create a discursive space (Dunne & Raby, 2013). The scenarios occur in the somewhat near future and help envision the interaction and experience with these technologies. Importantly, it does not concern what is currently technologically possible but rather opens the space for discussion and exploration. At the end of the design fiction, there was one instruction that would guide the sound designer to create a musical utterance that should be both legible and trigger an emotion.

3.2.3 Designed sounds for three scenarios

3.2.3.1 Scenario: obstruction

For the first scenario, people blocking the way of the robot, the sound designers had the idea of making a sweeping sound to imitate the movement of stepping aside. By adding chords to the sounds they aimed to give the sound a nicer and more friendly touch. In the first part, the sound communicates that something is wrong with a sound similar to soft beeping. At the end of the sound, staccato piano sounds aim to convey a nice but determined sound to move away. The idea of the sound designers was to keep the sound simple and friendly as the robot wanted something from the people while simultaneously conveying some tension as the sound has a relatively fast pace to convey urgency.

3.2.3.2 Scenario: closed door

For the closed door scenario, it was aimed that the sound be more alarming. The robot has to draw the attention of people passing by so that they may help it by opening the door. Therefore, the designers intended to make the beginning of the sound more attention-grabbing. This alarming part was enriched by adding more musical sounds. Additionally, the sound designers intended to imitate a knocking sound as if the robot is knocking on the door. This part was also determined to show some level of distress.

3.2.3.3 Scenario: robot stuck

For the sound of the robot getting stuck scenario, the sound designers tried to resemble some sort of stumbling or tripping sound. At the same time, they also wanted it to sound like a few steps were taken. Therefore, the sound started with two short different tones playing alternately. This is followed by a sequence of tones going down and being played more legato. The latter part was also intended to make the sequence sound more dramatic to enhance the problem being conveyed.

3.2.4 Video design

All three scenarios were enacted using the Kuka IDo robot (see Figure 1), as the IDM Harmony robot was not on location at the time of the recording. We selected a hallway at the University of Twente that looked similar to a hallway at a Dutch hospital we visited earlier. The robot enacted each of the scenarios and actors were instructed to play as hospital staff, but not respond to the robot. Once the robot required help, it would play one of the musical utterances, and the recording then stopped.

Figure 1. Screenshots of the videos used in the experiment.

3.3 Method

3.3.1 Participants

The participants were recruited through the crowd-sourcing platform Prolific. All participants received $3,- for their participation. The inclusion criteria for the study were that the

participants had to be from the United States of America, not have any hearing or cognitive impairments, have completed 100 or more studies on Prolific, and have an approval rate of 95% or higher. In total, 199 participants participated in the study. From these, four participants were excluded because their completion time was very long or very short, they did not answer one or more of the attention checks correctly and their answers also gave reason to assume they did not pay close attention. Another seven participants were excluded because they did not pass several attention checks, which gives reason to assume they did not pay close attention which again was also indicated by their answers. This resulted in a total of 187 respondents (87 female, 99 male, and 1 non-binary) that were included in the data analysis. The mean age was 36.76 years (SD = 12.23).

The study was reviewed and approved by the ethics committee of the faculty of Electrical Engineering, Mathematics, and Computer Science of the University of Twente and registered under the reference number 230072.

3.3.2 Experiment Design

The videos were assessed through a video HRI study. The study was designed as a 1 by 3 between-participants design with random assignment of participants. The independent variable is the *sound condition* which has three levels: no sound, a simple beeping sound, or the use of musical utterance. The “no sound” condition serves as a baseline to see to what extent the situational context signals the robot's intent. We also included a condition where the robot would use a simple beeping sound to assess whether the complexity of music improves emotion elicitation or intent communication.

3.3.3 Procedure

SurveyMonkey was used as a platform for the questionnaire. Participants received a link through Prolific to access the questionnaire on SurveyMonkey. Participants received an introduction to the questionnaire and the use case at the beginning of the questionnaire. It was explained that the videos take place in a hospital and their tasks were described. After this, the participants answered the questions regarding their experience with robots. Then the participants watched the first video, after which they filled in several open and closed questions. This procedure was repeated for the remaining two videos. At the end of the questionnaire, the participants were asked some background questions. Before filling out the questionnaire, the participants received an information brochure and had to consent to participate. After the background questions, the participants were debriefed and were again asked to give their consent to ensure they still consent after having all the information after they received the full information.

3.3.4 Measures

3.3.4.1 Legibility

The measure of legibility was done similarly to Kim and Follmer (2021). The participants had to answer the open question “What do you think the robot is trying to communicate?” and rate their *confidence* in the answer on a 7-point Likert scale. The open question was coded based on the level of *accuracy* with which the participants described their answers. The open question was binary coded, where the value depended on whether the answer mentioned something related to the robot's problem. The inter-rater reliability was indicating substantial agreement ($\kappa = .77$).

3.3.4.2 Empathy-evoking emotions

To measure the recognition of emotions, we asked the open question ”What emotion is the robot displaying? (Think about emotions like happy, sad, disgusted, neutral, angry)”. This question was coded as neutral (0), positive (1), and negative (2) emotions. We decided to code the emotions in this way because negative emotions more strongly evoke empathy than positive emotions or a neutral state. The value therefore reflects the likelihood of evoking empathy. The inter-rater reliability was strong ($\kappa = .88$). In addition, we asked participants to indicate the intensity of the attributed emotions on a 7-point Likert scale.

3.3.4.3 Empathy

Empathy was measured by rating a set of statements related to situational empathy on a 7-point Likert scale. Here, 1 meant the statement was not true at all and 7 meant it was completely true. The statements were inspired by the statements of the Questionnaire of Cognitive and Affective Empathy (Reniers et al., 2011). Because this questionnaire is designed to measure the levels of *dispositional* (cognitive and affective) empathy, and we are interested in *situational* empathy, we adjusted the questionnaire accordingly. The combined factor analysis for all scenarios together showed excellent reliability. For cognitive empathy, Cronbach’s alpha was .93 and for affective empathy, it was .92.

3.3.4.4 Prosocial Behavior

To measure whether the participants intended to engage in prosocial behaviour they were asked the open question “What would you do if you were a person in the situation?”. The answers were coded as no prosocial behaviour (0), weak prosocial behaviour (1), and strong prosocial behaviour (2). No prosocial behaviour meant that the participant indicated that they would not do something to try to resolve the situation of the robot. Weak prosocial behaviour was behaviour related to resolving the problem but not resolving the problem. Lastly, strong prosocial behaviour meant that the participants would do anything they could do to resolve the situation. The inter-rater reliability was indicating a strong reliability ($\kappa = .83$).

3.4 Results

3.4.1 Legibility

An ordinal logistic mixed model was run to test whether the sound condition influenced whether the communication of the robot was legible or not. We found no evidence that the musical utterances were significantly different in legibility compared to the beeping condition (b = 0.16, t = 0.54, p = .590) or the no sound condition (b = -0.30, t = -1.13, p = .260). To analyse the participants' confidence in their interpretation of the robot for each of the conditions, we ran a linear mixed model with age and experience as covariates. For the musical utterances, participants were less confident in their interpretation of the robot's behaviour than for the beeping condition (b = 0.53, t = 3.35, p = .001). We found no difference in confidence when we compared the musical utterances with the no sound condition (b = -0.27, t = -1.69, p = .091). Participants' age (b = 0.02, t = 4.60, p = <.001) and experience with robots (b = 0.20, t = 3.53, p = <.001) did significantly influence their confidence.

3.4.2 Empathy-evoking emotions

We ran an ordinal logistic mixed model to test the influence of the sound conditions on empathy-evoking emotion with participants' experience with robots as a covariate. Musical utterances were significantly better at evoking emotions that can lead to empathy than a beeping sound (b = -0.40, t = -1.78, p = .048), but showed no difference between the no sound condition (b = -0.30, t = 1.48, p = .138). Participant's experience with robots had a small positive effect on evoking emotions that can lead to empathy (b = 0.02, t = 2.56, p = .037). We found no differences in terms of the intensity of the attributed emotions between the musical utterances and the beeping sound (b = -0.03, t = -0.17, p = .864), nor the no sound condition (b = 0.01, t = 0.05, p = .957).

3.4.3 Empathy

The influence of sound conditions on cognitive empathy was tested by running a linear mixed model including the participants' age as a covariate. The musical utterances were significantly better at evoking cognitive empathy compared to not using sound (b = -0.38, t = -2.18, p = .030), but showed no difference when compared to the beeping sounds (b = -0.10, t = -0.58, p = .562). Participants' age did not significantly influence cognitive empathy (b = -0.01, t = -1.01, p = .314).

For affective empathy, we also ran a linear mixed model and again included the participants' age as a covariate. We found no significant differences between musical utterances and the beeping sound (b = -0.20, t = -1.48, p = .140), or no sound (b = -0.14, t =

-1.04, p = .298), in terms of the affective empathy that was evoked. Age was also not significant (b = -0.01, t = -1.15, p = .250).

3.4.4 Prosocial behaviour

We ran an ordinal logistic mixed model to test the effect of sound on prosocial behaviour. We found no differences in evoking prosocial behaviour between the musical utterances and the beeping sound (b = -0.00, t = -0.01, p = .991), nor the no sound condition (b = 0.21, t = 1.08, p = .282).

3.5 Discussion

In this online video-HRI study we investigated how musical utterances influence the robot's legibility and ability to evoke emotions that may lead to empathy in bystanders. We compared these sounds to a condition where participants only had the context to make sense of the scenario and a condition where the robot would make a beeping sound to attract the attention of bystanders. We developed the musical utterances and beeping sound for the robot together with two sound designers. The sounds were assessed in terms of their legibility, ability to evoke emotions that can lead to empathy, the empathy that was evoked, and prosocial behaviour.

The musical utterances made it significantly easier for participants to interpret the robot's "experienced" emotions (cognitive empathy). However, we found no evidence that musical utterances improved participant's interpretation of the scenario compared to when the robot would use no sounds, nor that they were better at evoking emotions that can lead to empathy. When the robot used a beeping sound to draw attention, participants felt more confident in their interpretation of the scenario, even though this did not significantly improve their ability to interpret the scenario. However, the beeping sound was worse at evoking emotions that can lead to empathy than the other two conditions.

Overall, we found that participants could interpret the scene well regardless of the condition with ~75% of them coming up with the intended interpretation. Even though we pilot tested the ambiguity of the scene, it appeared that even in the condition with no sound, the situational context was often sufficiently clear to participants that they could infer the robot's intention from its non-verbal communication. This ceiling effect is problematic for our investigation, because there was little room for the sounds (whether they are beeping sounds or musical utterances) to improve legibility, making it difficult to study the effects of the sounds in resolving ambiguity. Our results are therefore inconclusive in terms of legibility. Note however that we also did not find any evidence that the musical utterances make the legibility worse than when no sound is used. In other words, they would communicate something else than the situational context (i.e., incongruent multi-modal behaviour).

To conclude, while musical utterances may be a good way to communicate emotions, they are difficult and laborious to develop. Even with the help of sound designers, it is difficult to design musical utterances that can communicate intent well and elicit emotions that can lead to people helping the robot better than a simple beeping sound. Because we found no strong evidence of the benefits of musical utterances over no sound or a beeping sound, we will be looking at different SFUs, such as gibberish speech or non-linguistic utterances, to communicate with for the IDM Harmony robot.

4. Emotion categorization of non-verbal utterances across age and match between different SFUs and the Harmony robot

4.1 Introduction

To advance our research on the design of IDMind's Harmony robot's voice, we aimed to investigate how users perceive and categorise emotions conveyed by the robot across various age groups. In healthcare settings, predominantly comprising older adults, it is crucial to explore how this demographic, which often has distinct needs, perceives and comprehends robot communication.

As mentioned in Section 3, one approach to robot communication involves using semantic-free utterances (SFUs). These sounds, devoid of semantic content or words, come in various types (e.g., gibberish speech, beep-like) with different levels of human likeness. Research emphasises the importance of aligning a robot's voice with its appearance, matching people's mental models of robotic sounds and looks. For instance, humanlike voices are considered more suitable for humanlike robots.

Gibberish speech is a type of SFU that comprises of human-like vocalisations that resemble human language but lack intelligible content (e.g., Simlish, Teletubbies)m while non-linguistic utterances (NLU) rely on acoustic cues, which may encompass a variety of sounds such as whirrs, boops, beeps, or other distinctive auditory signals (e.g., R2-D2). These two types of SFU have been developed and successfully evaluated in previous HRI studies to communicate robot emotional states. The Bremen Emotional Sound Toolkit (BEST) (Castellano et al., 2013), part of the EMOTE project, comprises synthetically generated acoustical emblems created by experts in music production, sound design, and emotion psychology. These sounds considered non-linguistic utterances (NLUs), encompass emotions and speech acts such as affirmation or encouragement. Emotions include anger, disgust, enjoyment, fear, sadness, surprise, interest, and shame. The EMOGIB (Yilmazyildiz et al., 2011) corpus, designed for robot-child communication and later validated with adults, features gibberish speech expressing six emotions: anger, disgust, fear, happiness, sadness, and surprise. An actress portrayed these emotions using scripts generated from specific syllable combinations, considering vowel and consonant distributions of the Dutch and English languages.

Research has shown that older adults have a decreased ability to recognize vocal emotion. Several hypotheses have been put forward to explain this effect on emotion recognition. The cognitive hypothesis suggests that age-related changes in neural structures, such as the deterioration of the frontal and temporal lobes, and the sensory hypothesis attributes poorer emotion recognition to age-induced sensory decline, such as hearing loss. Other research on prosody and semantics has shown that their relative importance in message decoding changes with age. Older adults appear to increasingly depend on semantic information. The aging effect appears to be emotion-specific as research shows a consistent trend in of some

emotions being easier to identify than others. Understanding how these age-related effects affect emotion categorization accuracy and vary for different emotions emphasises how crucial it is to carefully think about how we recognize emotions in various situations and with different emotional states for the design of age-inclusive robot communication.

In this study, we focus on advancing the design of age-inclusive SFU robot communication by enhancing the understanding of how age and its interaction with other user characteristics affect the emotion recognition of different robot SFUs. Additionally, we explore how people evaluate two types of SFUs and their suitability for the IDMind's Harmony robot task and appearance. To address these inquiries, we conducted a pre-study to narrow down the available stimuli that effectively communicate specific emotions and to assess participants' preferences regarding the compatibility of these SFUs with a healthcare robot. The results also guided the addition and design of a third database of robot SFUs. This database had qualities similar to language but maintained a distinct robot-like quality, distinguishing it from the other two SFU databases. The main study evaluated emotion recognition performance across age groups using the stimuli selected during the pre-study, along with those from the additional database.

4.2 Pre-Study

Because the BEST (NLU) and EMOGIB (GS) databases each contain multiple audio recordings of various emotions, we conducted a 2 (SFU database) x 5 (emotion category) within participants' online audio-based study study. The goal of this pre-study was to narrow down the available stimuli to those that most clearly communicate a specific emotion per database. Additionally, to inform the sound design and selection for IDMind's Harmony robot, we looked into what considerations participants make when matching this robot with one of two SFU databases.

4.2.1 Method

4.2.1.1 Participants

To participate in the pre-study, specific inclusion criteria were established. Participants needed the ability to hear sounds from their electronic device and had to be 18 years or older. The survey was conducted online to ensure that participants could hear the stimuli from the device they used for the survey. Convenience sampling was used to recruit participants, resulting in 24 participants (10 men, 14 women) consenting to take part. The sample primarily comprised young adults (mean age = 24.9, SD = 4.2, range: 19-34), with 23 participants identified as Dutch and 1 as German. Among them, 23 had a University education, while 1 had an education from a University of Applied Sciences. Participants' prior exposure to robots varied: 1 reported daily contact, 3 weekly, 5 monthly, 7 yearly, 7 had

limited interactions, and 1 had no prior contact with robots. This study obtained approval from the Ethics Committee in Computer & Information Science at the University of Twente.

4.2.1.2 Stimuli

The stimuli are sourced from the EMOGIB and BEST databases, each housing distinct emotion categories. To ensure an equitable comparison between the databases (GS vs NLU), only emotions with identical names were matched: anger, disgust, fear, sadness, and surprise. We selected eight random audio files for each emotion in each database, resulting in a total of 80 recordings (40 per database). From EMOGIB, we randomly chose 8 stimuli per emotion category from the first two Dutch-based sub-corpora, aligning with our Dutch-speaking target demographic from the Netherlands. In the BEST database, we picked 4 low and 4 high emotional intensity stimuli per emotion category to reach a total of 40 stimuli from each database.

4.2.1.2 Procedure

The study, conducted via Qualtrics, commenced with participants consenting to data use. They adjusted device volume using sample audio files before engaging with 5 blocks, each featuring 16 items (8 audio files multiplied by 2 SFU types) representing intended emotions. Participants rated emotion clarity on a 6-point Likert scale and provided reasoning after each block.

Additionally, the study explored the suitability of two SFU databases for IDMind's Harmony robot. After the audio blocks, context about the robot was provided; that it is a healthcare robot designed to deliver biological samples between laboratories inside the hospital. Participants listened to randomly chosen audio samples from each dataset and identified which set they thought the robot would produce. They explained their choice, first without visuals and then with a picture of the Harmony robot for size comparison. Demographic data like age, gender, nationality, education level, mother tongue, and participants' exposure to robots were then also collected.

4.2.2 Results and Discussion

There is variation in how clearly emotions were conveyed among stimuli, across emotions, and datasets. Overall, surprise was most clearly conveyed, followed by sadness, fear, disgust and anger. On average, the intended emotions were more clearly conveyed with EMOGIB than with BEST sounds. We selected for each emotion and each database three audio stimuli that most clearly conveyed the intended emotion, which yielded 30 selected audio stimuli.

We also asked participants to associate one of the two SFU databases with IDMind's Harmony robot. Because we were interested in the influence of the robot embodiment on the participants' choice, information about the robot's role and work environment was provided to participants twice: once without a picture of the robot and once with a picture

of the robot. When no picture was provided, 79% of participants matched the robot with the BEST sounds; this percentage increased to 92% when a picture of the Harmony robot was included.

We also conducted a thematic analysis on the open-ended questions that prompted participants to elaborate on their choice of robot sound to match with the Harmony robot, four themes emerged from participants' perceptions of suitable sounds for a hospital sample delivery robot.

1. *Gibberish speech and language-likeness:* Most participants reported that because EMOGIB sounded too much like a language, it confused them as it was not understandable. However, a small number of participants found that EMOGIB's language-like quality allowed for clearer communication.
2. *Vocalizations matched the robot:* Participants indicated that their choice of the database was influenced by how closely the vocalisations aligned with their personal expectations or mental model of what a robot should sound like, the tasks the robot would perform, and the environment in which it would operate.
3. *Need for a clear distinction between robot and human vocalisations:* Participants emphasised the importance of a clear distinction between robot and human vocalisations. They preferred the sounds from the BEST corpus because it signified that the sounds originated from a robot, not a human.
4. *Gibberish is more expressive/emotional:* Some participants reported that EMOGIB was better at conveying emotion. However, half of the participants who reported this viewed it as a positive characteristic, while the others regarded it as an undesirable characteristic for robots.

Participants selecting between the two databases while viewing the Harmony robot's picture increasingly mentioned being influenced by the matching qualities of the robot and the utterances. This led to expanding the subtheme of *Vocalisations that matched the robot* to include considerations related to the robot's *appearance* and its *perceived capabilities for producing the sounds.*

Additionally, participants reported emotional attributes that affected their vocalisation preferences for the robot, encompassing perceptions of confidence, complexity, anxiety, irritation, annoyance, childlike qualities, drama, and high pitch for both databases.

As judged by the participants, sounds from the EMOGIB more clearly conveyed a specific emotion. While this could be an argument for designers to use EMOGIB sounds for expressive robot communication, at the same time participants judged EMOGIB less appropriate for a robot like IDMind’s Harmony robot. They mentioned that human-like vocalisations did not align with the robot's task or their expectations of how a robot should sound, and they expressed a need for a clear distinction between sounds coming from a

robot or a human. These observations led to the creation of a third database called ROBOGIB, a robotized version of EMOGIB, that contains manipulated versions of the selected EMOGIB stimuli to sound more robotic while maintaining the same expressive quality of EMOGIB. The 45 selected audio stimuli from 3 different databases and 5 emotions will be evaluated in the main study.

4.3 Main Study

To test our hypotheses, we conducted a within-participants online audio study, using a 3 (SFU database) x 5 (emotion category) setup. Participants categorised emotions, listening to prestudy-selected audio stimuli along with additions from the third database (ROBOGIB). We explored the influence of age, gender, prior robot exposure, emotion category (including anger, sadness, disgust, fear, and surprise), SFU database (EMOGIB, ROBOGIB, BEST), and the perceived human-likeness of the sounds on participants' emotion categorization accuracy. Additionally, we also asked participants to rate the match between each SFU database and the Harmony robot based on its appearance, role (healthcare delivery robot), and environment (hospital).

4.3.1 Method

4.3.1.1 Participants

A total of 76 individuals took the survey, but 10 were excluded due to reported audio issues. Among the remaining 66 participants, ages ranged from 18 to 89 (M = 45.32, SD = 17.60), comprising 45 women, 19 men, and 2 non-binary individuals. Regarding prior robot exposure, 2 participants had daily contact, 2 weekly, 10 monthly, 17 yearly, 26 limited interactions, and 9 had no previous contact with robots. To participate, individuals had to meet specific criteria: being 18 or older, capable of hearing sounds from a device, and not self-reporting (mild) cognitive impairment. This experiment received approval from the Ethics Committee in Computer & Information Science at the University of Twente.

4.3.1.2 ROBOGIB Stimuli Design

In addition to EMOGIB and BEST, we created a hybrid set called ROBOGIB by modifying the original EMOGIB sounds to have a more robotic quality while preserving their expressiveness. Following principles from Wilson et al., 2017, we adjusted the gibberish stimuli by raising their pitch by two semitones and then overlaid this modified version onto the original stimuli with a 50ms delay. This manipulation was done using Audacity software. Consequently, this third set included the same 15 sounds as the EMOGIB set but in an altered form. Including this set expanded the total number of stimuli to 45, encompassing the top 3 sounds from 5 emotions and 3 databases.

The manipulated set of stimuli (ROBOGIB) hasn't undergone validation to assess its effectiveness in expressing intended emotions. It's unknown if the manipulation impacted the samples' expressive value. Despite this, we included the dataset as it might bridge the gap between the clear emotional expressiveness of EMOGIB and the suitable robot sounds of BEST.

4.3.1.3 Procedure

The study was held online via Qualtrics. Digital consent was given anonymously, followed by presenting three audio files that could be played to set the volume to a comfortable level. Participants then were presented with a total of 45 audio stimuli (3 audio files x 5 emotions x 3 databases) in a random order, each accompanied by an emotion categorization task and a human likeness rating scale. Participants were instructed to listen to the audio file and select the emotion that was conveyed. Choices consisted of anger, disgust, fear, sadness, surprise, and `other' in which participants could use any text to describe their choice of emotion. They rated the perceived human likeness of the audio file on a scale from ``1 = highly robot-like'' to ``6 = very human-like''. In the second part of the survey, participants answered some questions about how sounds from the 3 different databases match to the task, environment and appearance of IDMind's Harmony robot and the likeability of the voice (this part of the survey however is not further analysed in this paper). Finally, participant characteristics such as age, gender, nationality, education level, mother tongue and robot exposure (6-point Likert scale ranging from 1=daily to 6=never) were collected.

4.3.2 Results

4.3.2.1 Manipulation check

To assess if our ROBOGIB dataset achieved the goal of being less humanlike than EMOGIB, we used a linear mixed model analysis. This model took human-likeness ratings as the outcome variable, considering the database as a fixed effect and including participant ID and item as random effects. The analysis revealed a significant effect of the database on perceived human likeness ($F(2, 42) = 1135.57$, $p < 0.001$). Specifically, the first comparison showed that ALLGIB (including EMOGIB and ROBOGIB) was significantly more humanlike than BEST ($\beta = 2.404$). Additionally, the second comparison highlighted that ROBOGIB was notably less humanlike than EMOGIB ($\beta = -0.925$).

4.3.2.2 Effects

Age Effect. A significant relationship emerged between age and accurate emotion classification ($F(1, 2948) = 23.396$, $p < 0.001$). Older participants exhibited less accurate emotion classification ($\beta = -0.028$).

Effect of Emotion. An important effect of emotion on accurate emotion classification emerged ($F(4, 2948)$, $p < 0.001$). Comparatively, fear ($\beta = 1.113$), sadness ($\beta = 1.932$), and

surprise (β = 2.890) were better recognized than anger. Surprise had the highest recognition, followed by sadness, fear, disgust, and anger.

Effect of Database. A significant effect of the SFU database on accurate emotion classification was found ($F(2, 2948) = 26.854$, $p < 0.001$). ALLGIB (EMOGIB and ROBOGIB combined) had notably higher accurate recognition than BEST (β = 0.011). Also, ROBOGIB showed lower recognition compared to EMOGIB (β = 0.010). Emotion categorization accuracies were 69.5%, 62.1%, and 44.6% for EMOGIB, ROBOGIB, and BEST respectively.

Effect of Human-likeness. No significant effect of human-likeness ratings on emotion categorization accuracy was observed ($F(1, 2948) = 1.160$, $p = 0.292$).

Effect of Gender. Gender categories showed no variation in emotion categorization accuracy ($F(2, 2948) = 2.263$, $p = 0.132$). Although the glmer output indicates differences between non-binary and female participants (reference level), the uneven distribution of gender categories in our sample needs consideration (there were only two non-binary participants).

Effect of Robot Exposure. There is no significant main ($F(1, 2948)=.715$, $p = .367$) or simple effect of robot exposure on emotion categorization accuracy.

Age x Emotion Interaction. A significant interaction emerged between age and emotion concerning emotion categorization accuracy ($F(4, 2948) = 6.038$, $p < 0.001$). Analysing simple effects reveals a smaller age effect for recognizing fear (β = 0.016) compared to anger recognition. However, for sadness recognition, the age effect was more pronounced (β = -0.022). See Fig~\ref{fig:plot_age_emotion} for details.

Age x Database. The interaction effect between age and SFU type appears significant when considering the database with three levels ($F(2, 2948) = 3.347$, $p < 0.05$) in the afex version of our glmer model. However, this interaction seems driven by age effects in ROBOGIB differing from EMOGIB and BEST databases, which isn't a reliable comparison. The interaction between age and database isn't strongly supported by either of the two (Helmert-coded) SFU contrasts in our original glmer model.

Interaction effect of SFU Database x Robot Exposure. The interaction between SFU database and robot exposure is not significant ($F(2, 2948)= .125$, $p = 0.885$).

4.3.2.3 SFU databases match across robot appearance, role, and environment

A rapid mean comparison analysis of the participants' ratings concerning the alignment between the three distinct SFUs and IDMind's Harmony robot, spanning its appearance, role, and environment, indicates a preference for the BEST database concerning the robot's appearance and role. However, participants displayed a preference for the EMOGIB database

when considering its usage within a hospital environment, attributing higher likability ratings to this database for the robot.

For the open questions, there were several themes identified in regards to what aspect of the sounds influenced participants' rating of its match across the environment, role, and appearance.

Robot- and human- likeness. Some participants preferred sounds that were more robotic, as they felt it better suited the robot's appearance and role, while others preferred a more human-like quality for clearer communication, which was important to them in a hospital environment.

Match with expectations for robot sounds. Participants reported to be influenced by their mental models of what robots should sound like and if they expect the voice they heard to come from a robot.

Hierarchical Influence of Robot's Characteristics on Sound Evaluation. Participants tended to prioritise between these characteristics over the others, with more focus on the robot's role/task and appearance rather than the environment.

Appropriateness and fit of emotional load. Participants reported to be influenced by their perception of the emotional content of the sounds and their fit for robots in general, or the Harmony robot and its role. Participants found the emotional content on the sounds to be unnecessary, especially in environments like hospitals where a more neutral tone might be preferred.

Clarity of communication. Participants evaluated the sounds based on their perceived ability to clearly communicate with users around the robot. Clear communication was considered important, particularly in environments where effective communication is crucial, such as hospitals.

Sound qualities. Participants are influenced by various sound qualities such as pitch, intonation, and intensity. These qualities contribute to participants' perception of the sounds and their suitability for the robot's role, environment, and appearance.

In summary, participants' ratings of sound match across the environment, role, and appearance of the robot were influenced by the fit with the robot's characteristics, preferences for robot- or human-likeness, expectations for robot sounds, appropriateness of emotional load, clarity of communication, and various sound qualities.

4.4 Discussion

This study explored the potential factors impacting the emotion categorization accuracy of SFUs, including age, gender, robot exposure, emotion type, SFU database, and perceived

human likeness of SFUs. Consistent with prior research, our findings reveal that emotion categorization accuracy tends to decline with age and is influenced by both emotion type and SFU type. These findings highlight the complex interplay between age, emotion type, and the choice of SFU in emotion recognition accuracy of SFU. Additionally, we probed participants to choose a set of sounds for IDMind's Harmony robot between two databases and explained their matching preferences. We found three main factors that influenced their decision: an antipathy for language-like sounds, a need for a robot sound distinctive from human voices, and their mental model of how robots should sound. Understanding these dynamics between listeners' emotion recognition accuracy and robot voice preferences can inform the design of more inclusive and effective robots that consider age-related differences, prioritise relevant emotion types, and select appropriate SFUs for robots. Furthermore, it underscores the necessity for further research, delving deeper into these interactions across a wider spectrum of emotional stimuli to develop more comprehensive insights for future robot designs in healthcare settings. Additionally, we were interested in identifying which SFU was a better match for the Harmony robots across its appearance, role, and task. We found that while participants rated the BEST database as a better match for the robot's appearance and role, they preferred the EMOGIB for the environment (hospital) the robot would be in. When asked which aspects of the sounds influenced this decision, participants reported robotlikeness, prior expectations, sound qualities (i.e., pitch), clarity, prioritisation between robot characteristics (role, environment, and appearance), and emotional load.

5. Design of Robot Movement Behaviours with Performers

5.1 Introduction

Robots present in public spaces are bound to make mistakes. To ensure the successful navigation of the Harmony robot within a hospital environment and its acceptance by people, it is crucial for the robot to respond appropriately to these errors. These errors can be classified into two categories: social errors and performance errors (Tian and Oviatt, 2021). Social errors refer to mistakes that undermine the user's perception of the robot's socio-affective competence, while performance errors erode the user's expectations of the robot's intelligence and competence (Tian and Oviatt, 2021). If robots fail to address these mistakes in a manner deemed appropriate by the users, it can result in a loss of trust in the robot's capabilities and potentially lead to the abandonment of the technology (Kwon, 2016).

Previous research in the field of Human-Robot Interaction (HRI) has explored strategies to mitigate the negative effects of encountering error scenarios. These strategies include offering apologies, providing compensation, and giving forewarnings. In this study, we aim to leverage the available modalities of the Harmony robot to facilitate its recovery from these errors. Drawing from sociological theories by Goffman (1967) and Schonbach (1980) on the phases and taxonomy of accountability, we will focus on communicating different types of apologies. Since the Harmony robot does not employ verbal speech, this investigation centres on how the robot can utilise *movement* to effectively convey various types of apology strategies for social mistakes in a healthcare environment.

To address this question, we have designed a study that aims to explore and observe different ways in which the Harmony robot can utilise movement to mitigate the negative impact of social errors, including invasion of personal space, interruption of hospital calmness, and lack of social behaviours. We will observe how movement experts, such as actors and dancers, approach these error situations when performing as the Harmony robot. By studying their techniques and gaining insights, we can acquire valuable knowledge on how movement can be effectively utilised to communicate apologies and facilitate positive interactions in a healthcare setting. The findings will be used to *create universal descriptions* of these behaviours, specifically describing the performer's movements in a way that they can be transferred to robots. These descriptions will then be applied to the specific behaviours of the IDM Harmony robot.

5.2 Method

5.2.1 Participants

Twenty-six participants were recruited from three distinct performative arts backgrounds: improv theatre (7), hip-hop dance (11), and modern dance (8). The age range of the participants ranged from 18 to 39 years (M=23.73, SD=4.68). Among the participants, fourteen identified as female, eleven as male, and one preferred not to disclose their gender. In terms of nationality, thirteen identified as Dutch, three as German, two each as Czech and Latvian, and one each as Estonian, Greek, Italian, Japanese, Mexican, and having a dual Russian/Latvian nationality. Regarding prior interaction with robots, fifteen participants indicated "No", ten responded "Yes", and one participant marked both "Yes" and "No".

The study was conducted in English, and all participants needed to be proficient in speaking and understanding the language, as well as being at least 18 years old. Additionally, participants need to be affiliated with the student associations we are collaborating with. As a token of appreciation for their participation in the main study and recognizing the use of their practice time, participants receive a €25 bol.com gift card and are provided with drinks and snacks during the sessions.

For the pilot study, we recruited five participants (4 males and 1 female) from the Human-Media Interaction group at the University of Twente. These participants have experience in various forms of performing arts, and research backgrounds in research and technology. The study received ethical committee approval (Reference number: 230360) and is being conducted at the University of Twente in the Netherlands.

5.2.2 Experiment Design

We designed a three-hour qualitative study aimed at designing creative and easily understandable robot movements when dealing with social errors. To achieve this, we collaborated with movement experts such as dancers and improv actors. We tasked them with portraying both the Harmony robot and hospital stakeholders, acting out three distinct social error scenarios that the robot might encounter while navigating the corridors. These scenarios included a navigation mistake, interrupting the hospital's calmness, and not meeting people's social expectations. To control the expressive range of the performers when portraying the robot's role, we imposed limitations on their modalities, such as allowing them to use only one arm or restricting changes to their shape and size. Additionally, we are also interested in looking at how the performers make use of the robot's sound modalities when communicating with the "stakeholders," as this could inspire future studies on the IDM Harmony robot's sound.

5.2.3 Measures

By observing these interactions, we aim to identify and annotate potential expressive movements that the robot can employ in similar situations by describing them in abstract format (i.e., robot independent) and using Laban Movement Analysis (Groff, 1995) which would allow us to translate the movements into universal descriptions and to the IDM robot and then test their acceptability across different contexts. Furthermore, we will closely observe how the performers utilise sound modalities to inspire future research on the robot's voice. This observation will help us gain insights into how different sound elements can contribute to the overall effectiveness of the robot's communication.

5.2.4 Procedure

Each study session lasted three hours. The first 30 minutes consisted of introductions, going over the consent forms, and providing a short summary of robots and their types. The next two hours were divided into two blocks, which were alternated between sessions.

Block A focused on studying performers who acted out error scenarios as the IDM Harmony robot and engaged hospital stakeholders to inform the design of error mitigation strategies related to the robot's movement. The objective of this block was to gain insights into how performers would imagine and anticipate the Harmony robot's error recovery behaviours by assuming the role of the robot. Their task was to enact three social mistake scenarios (Described in Section 4.2.4.2) that the IDM Harmony robot could potentially encounter while traversing a hospital corridor, either in the role of hospital stakeholders or as the IDM Harmony robot itself. Before the enactment, we created a fictional corridor by placing two ropes on the floor. The participants were divided into two groups: one group consisted of hospital stakeholders, and the second group consisted of the robot performer. For the robot performer, we provided it with instructions on their available modalities and the errors they would encounter or engage in and explained their roles to the stakeholder performers.

Once the participants were fully instructed and had no questions about their roles and tasks, we initiated the enactment. The first performer, taking on the role of the robot, acted out the error scenario three times. They were given the option to either repeat the same action or try something different after each round, aiming to enhance their comfort level in acting like a robot and encourage creative freedom. After the three iterations, we asked the robot performers how they utilised each modality. Subsequently, we proceeded to enact the next two error scenarios each time with a different robot performer. Upon completing the three error scenarios, we conducted a brief group interview and asked all participants about their feelings during the interaction, their experience while acting as the robot, and if they would approach anything differently.

Block B involved an improvisational session between the performers and the Pepper robot. They improvised with the robot and later pretended to be the robot itself. Block B was conducted by a fellow researcher who had her own research questions. With the aim of

studying how the relationship between performers and robots evolves during and after improvised interactions. However, if Block B was conducted as the first block, it could potentially affect the performers' familiarity when acting out as a robot.

Lastly, the last 30 minutes of each session comprised a round table exit interview. We asked participants about their experience during the study, what aspects worked or did not work for them, and provided a debriefing. Additionally, we allowed participants to ask any questions they had.

5.2.4.1 Introducing the Harmony Robot with VR

To reduce the amount of travel the IDM robot would need to go through to use it for this study and to not hinder its further development, we decided to introduce the IDM Harmony robot to the participants in a virtual environment using an Oculus Quest 1. The virtual environment runs in Unity and features a hospital corridor Unity asset[1] with a virtual version of the IDM Harmony robot situated inside. Participants could walk around the robot to experience its appearance and size. Using the controllers, they were able to let the robot drive along a predetermined path through the corridor, change the brightness, colours, and patterns of the LED lights, and cause the robotic arm on the robot's back to move. They could also cycle through different content on the robot's head display, such as images and different sets of eyes (Figure 2). The virtual environment also included hospital ambient noise to enhance the participants' immersion. Additionally, mechanical noises originating from the robot's navigation movement were incorporated. The IDM team provided this sound by recording the noise that the robot makes while navigating through a hall.

Figure 2. IDM Harmony robot navigating through the hospital corridor displaying different uses of modalities.

[1] Hospital Corridor by Bright Vision Games
https://assetstore.unity.com/packages/3d/environments/hospital-corridor-231428

5.2.4.2 Error Scenarios

We provided each robot participant with a description of each error scenario. These scenarios were inspired by an internal brainstorming session and discussions among various partners in the Harmony project. The scenarios are described as follows:

1. **Navigation error:** Picture the Harmony robot. Now imagine you are it. Your proximity sensors are not calibrated correctly so you invade an individual's personal space by getting too close to them. You can make use of your given modalities to respond, communicate, and signal them.
2. **An interruption of the hospital's calmness:** Imagine the Harmony robot. Now, picture yourself as the robot. While you are driving through the corridor you will make a social mistake by disturbing the peaceful and calm atmosphere of the hospital. You can do this in any way you want with your given modalities.
3. **Social expectation failure:** Picture the Harmony robot. Now imagine you are it. You have one minute to move from point A to point B. Remember you are carrying laboratory samples from patients, so the sooner you arrive at point B, the better. People will try to engage with you, but you have no time. You can use your given modalities to respond, communicate, and signal people around you.

Although the main description of the error scenarios was provided to the performers, we allowed them the freedom to imagine how these would take place. This information also helps us come up with different robot behaviours that can be deemed inappropriate by stakeholders.

5.2.4.3 Robot Modalities

We also provided the participants with four different modalities they could use to communicate as the IDM Harmony robot. These were a robot arm, whole-body movements, non-verbal sounds, and changes of shapes and size.

5.3 Pilot Results

Because the pilot study was conducted with expert researchers, the focus of the pilot test was very much on improving the study design and we had to reduce the duration of it to a two-hour session. Based on their feedback, we adapted the study’s instructions to be clear enough for the performers to follow and allow them to be comfortable and creative during the sessions. The main suggestion that was made was to involve the performers in the study and to let them know what we need from them and point out when they do something wrong or right while enacting the scenarios and more appropriate points to conduct short interviews and the plenary questions. These changes are now integrated into the research procedure outlined in this deliverable (Section 4.2.2).

During the pilot test, the first robot performer enacted a scenario involving a social expectation failure, while also having the ability to change its shape. In this scenario, a stakeholder approached the robot closely and deliberately interrupted its journey by placing their leg in front of the robot to block its path. The stakeholder then moved their leg away and back in front of the robot, blocking the path as if testing its sensors (Figure 3). Although the robot performer did not utilise the modality of changing its shape or size, it stopped its navigation and employed vocalisations throughout the scenario, specifically "boop" sounds. The robot performer adjusted the pitch, intensity, and duration of the boop sounds to convey different states or responses. For example, when the stakeholder interrupted the robot's navigation, the robot performer emitted a boop sound with a higher pitch and longer duration, pausing its constant boop sounds while moving in the corridor. When the stakeholder moved away, the robot performer resumed its constant booping sound and continued navigating the corridor. Furthermore, the robot maintained its orientation towards the other side of the corridor and did not execute any turns or rotations during this particular scenario.

Figure 3. The performer acting as a hospital visitor (purple) interrupted the robot’s navigation by placing their leg in front of it. In response, the robot performer (blue), manipulated its vocalisations.

For the next scenario, the same robot performer assumed the role of the Harmony robot. They re-enacted the same error scenario, but this time, they possessed the additional

modality of using one arm instead of being able to change shape and size. Throughout this enactment, the robot performer continued to emit sounds, resembling a constant humming sound similar to the sound of "u," while portraying the robot. As stakeholders approached the robot, they attempted to engage in social interactions by obstructing its path and initiating conversation. To address these situations, the robot performer initially attempted to navigate around the stakeholders. If the stakeholders persisted in following the robot, the performer ceased navigation, raised their arm to eye level with the palm facing the stakeholders, and emitted an alarm-like sound. In Laban's terms, the performer stopped its navigation, stood straight, and moved their arm and hand in a lightweight, sustained in time, moves directly in space, and has a bound flow.

The constant humming ceased only when both navigation and arm movements stopped, resuming when the robot resumed navigation or raised its arm (Figure 4). When the first participant (Stakeholder 1) engaged in this interaction, they moved away from the robot's path. However, the subsequent stakeholders (Stakeholders 2, 3, and 4) who engaged with the robot by surrounding it did not move away after the robot repeated the same strategy. It was only when Stakeholder 1 demonstrated to the other stakeholders that the robot would perform these gestures when its path was blocked that the rest of the stakeholders avoided obstructing the robot's way.

Figure 4. Robot performer (blue) raising their arm when approached to engage in a social interaction.

The second performer portrayed the scenario of interrupting the calmness of the hospital while possessing the additional modality of changing shape and size. However, during this enactment, the performer did not utilise the extra modality and resorted to producing loud alarm-like sounds while remaining stationary. As two stakeholders (Stakeholder 1 and 2) began to approach, they exhibited curiosity, attempting to ascertain the reason behind the robot's alarming noise. Then another stakeholder (Stakeholder 3) of the stakeholders acted out as if they were taking selfies, pictures, and videos of the robot making this sound, and Stakeholder 4 walked by the robot while covering their ears. Subsequently, the robot performer initiated slow movement through the corridor while continuously emitting the alarm sound (Figure 5).

Figure 5. While the robot performer (blue) was making loud alarm sounds. A performer portraying a visitor (light purple) simulated recording the robot with their phone, while the other visitor (dark purple) covered their ears as they approached the robot. The staff performer (green) observed the interaction and expressed discontentment regarding the noise.

5.5 Main Study

To inform the design of different movement behaviours the Harmony robot can employ when it makes a social error, we set up six sessions where different performers from different performative art backgrounds performed as the IDM Harmony robot or as hospital users (visitors or staff). When we looked at the videos for this study, we looked at the errors the participants made when they were acting as the IDM Robot, the response from the participants acting as the users around it, and again at the response of the participants acting as the IDM robot to the users.

5.5.1 Robot Performers' Error Behaviours

When participants were prompted to make navigation errors, these were the most common ones:

During navigation errors performers mostly engaged in violating proxemic behaviours such as bumping into a user (4 times) or a wall, (5 times) and invading their personal space (22 times), some while also using the robot arm (20 times). Additionally, they also followed users around to get close to them (5 times).

When participants were asked to enact errors that disturbed the peace of the hospital environment they usually engaged in proxemic errors again or when they were able to use sounds (9 times), engaged in making loud random noises, some of them while stopping in the middle of the corridor (2 times), and used their shapeshifting behaviours to stretch themselves to the side by extending their arms to block the user's way (3 times).

For the social errors, we were interested in seeing how the participants acting as the Harmony robot dealt with users' interruptions to get to point B. When the participants were able to use an arm they pointed towards the next position (18 times) it had to go to. Meanwhile, when participants were able to use sound, they used it to communicate with people to move out of the way (36 times).

5.5.2 User's reactions to Robot Errors

When the robot invaded the user's personal space, the user's performers attempted to stop the robot from approaching them (28 times) by grabbing it and directing it in another direction (24 times), grabbing it and halting its movements (3 times), preventing its forward movement by stretching their arm to impede its progress (2 times), moving the robot to a different position (3 times), and hugging the robot once.

When the robot displayed shapeshifting behaviours to disrupt the hospital's calmness, they attempted to manipulate its shape once. Meanwhile, when the robot used sounds to disrupt the calmness, they tried to stop it by 'pressing its buttons' (4 times).

When participants were asked to be curious about the robot while it was delivering items in the hospital's corridor, they blocked the robot's way (36 times), pretended to take selfies with the robot (2 times), and inspected the robot (3 times).

5.5.3 Robot's Mitigation Strategies

During the study session, not a lot of the robot performers made use of mitigation strategies. However, when they did, they got away from the user they offended (8 times), bowed towards them (7 times), spun around (3 times) or tilted their head (1 time) to show confusion, and they signalled "sorry" in sign language (1 time).

5.6 Discussion

This study focuses on designing robot movement behaviours within healthcare contexts and social scenarios involving errors. The research team collaborates with movement experts and employs robot performers to enact diverse social error scenarios. Through pilot testing and the main study, valuable data are collected on how the robot performers use different modalities to engage stakeholders, elicit responses from participants, and mitigate the negative impacts of encountering social errors. The study aims to comprehend interaction dynamics and pinpoint patterns in robot modality use, particularly in movements. The findings will aid in refining robot behaviour designs to ensure they facilitate meaningful and engaging interactions while minimising the adverse effects of social errors.

Despite the pilot study primarily centring on discussions and refining the study's design, we gathered significant data on how the robot performer utilised available modalities to engage stakeholders. The pilot study provided insights into the robot's modality use and interaction dynamics between performers and participants acting as stakeholders, shedding light on participant reactions to simulated robot behaviours. Meanwhile, the main study exhibited various behaviours that both users and the robot could enact when encountering and simulating robot errors. This facilitated identifying people’s expectations regarding the robot's behaviour. Additionally, the unspecified nature of described error scenarios allowed participants to simulate various social errors, offering opportunities to learn about potential errors the robot might encounter while navigating hospital corridors.

6. IDM Harmony robot behaviour design for the Harmony use-cases

6.1 Introduction

The studies outlined in Sections 3 and 4 relate to specific aspects of the Harmony robot's behaviour where we want to push beyond the state-of-the-art; developing the robot's voice, robot behaviour that is specifically aimed at mitigating the negative effects of a robot making a mistake, and adapting the robot's behaviour based on the situation context (e.g., expressing sorry in a hallway versus in a lab setting). In addition, we are also developing robot behaviour that the robot will need to address interactions with people it can encounter in the hospital hallways. The study that is scheduled for the end of September at the University Hospital in Zurich will also evaluate these behaviours. In this Section, we report on what these behaviours are and how we aim to use them.

The objective of behaviour design is to create behaviours that people can understand, communicate pleasantly, and effectively achieve their goals. To achieve this, the IDM Harmony robot has multiple ways to communicate with people. It can use sound, produced by two speakers, as well as lights located in its head and base, and it can express itself through motion and gaze. Furthermore, the tablet that forms the robot's head can display different types of information. By default, the tablet shows the robot's eyes.

The behaviour of the robot will vary depending on the specific situational context within the hospital. For example, when faced with an event where people show interest in the robot and block its path, we aim to differ the way urgency is communicated depending on in what "contextual zone" (e.g., hospital hallway or the laboratory) the event occurs. This adaptation of robot behaviour is important because the acceptable method of resolving the event may vary per zone. In the laboratory, where personnel need to concentrate on their work, distracting sounds would not be welcomed. Thus we would look at communicating through different modalities that do not include sound. On the other hand, in the hospital hallways, the use of sound may be more acceptable. To endow the IDM Harmony robot with this adaptive behaviour capability, we have implemented and tested different contextual zones that allow for specific behaviour policies to be set. For instance, these zones include a silent zone where the robot's volume is reduced. These contextual zones can be marked on the robot's map of the environment.

In the next section, we will first provide an overview of these events, as well as draft strategies for the HRI component to respond or adapt to these events. One outcome of this exercise was to understand where and how certain robot modalities could be of use. This then informed our approach for the first integration of the HRI component on the IDM Harmony robot performed in June, which we discuss in the section after.

6.2 Hospital hallway events

6.2.1 Identified events

To approach the design of the robot's behaviours, UT initially took stock of the instances where the robot would need to interact with people during the delivery of goods within the hospital. By considering the events identified in deliverable D1.2, drawing from our own experience with the IDM Harmony robot in the hospital, and engaging in an internal brainstorming and discussion with other partners, UT compiled a preliminary list of events that the robot should be capable of handling. As we test the robot further inside the hospital, this list will be updated. Note that the behaviours to resolve each event can fail, in which case we would engage in mitigating the negative effects thereof based on the behaviours designed in the study described in Section 4. Also, note that whether we can address each of these events will also be determined by the hardware and software restrictions of the IDM Harmony robot.

We identified various ways the robot’s path may be blocked (obstruction event). The different obstruction events we have identified mainly differ in whether or not a human is blocking the way (instead of a static object), and how many, if any, other humans are present. Depending on this situation, we have drafted strategies trying to free the robot, or trying to get the bystanders to clear the desired way. Several events have to do with navigating. Either signalling behaviours to indicate the direction of the robot. Here the goal is to not startle people and help them anticipate the robot's trajectory. Other identified events include people who start to interact with the robot (e.g., out of interest), various events around entering, staying or trying to leave an elevator, and failure events such as the robot losing localization entirely or getting physically stuck. Finally, there are several Harmony-specific events, such as *ready to load*, *ready to start delivery* or *ready to unload*. All the identified events and related response strategies drafted for resolving these events are outlined in Appendix A.

6.2.2 Responding to events

To effectively respond to events, we distinguish between two types of robot behaviours: interactive and supportive. Interactive behaviours are focused on resolving an event by engaging with people. For example, if the robot's path is blocked, it could signal that it requires assistance to continue its primary task of delivering test samples. From a software perspective, the HRI module, responsible for determining which communicative behaviours to execute, takes control of the robot's operation and carries out the interactive behaviours until the event is resolved.

On the other hand, supportive behaviours are executed alongside the robot's ongoing operations. These behaviours are generally not directed towards a specific individual or intended to initiate an interaction. Instead, they aim to communicate with people in proximity to the robot. In these cases, not all communicative modalities of the robot may be available. For instance, while the robot is navigating, its movement cannot be used for communication. Instead, the robot can utilise LEDs to indicate the direction it is heading. During robot operation, specific behaviours may be activated to increase people's awareness of the robot (e.g., emitting a sound when making a turn) or enhance their understanding of the robot's intentions (e.g., blinking before a turn). Another instance of supportive behaviour is where the ongoing behaviour is altered. This could be alternations based on the contextual zones, but can also relate to certain encounters that do not require direct interaction. For instance, when the robot recognises someone has difficulty walking (e.g., by detecting crutches), the robot may slow down until it passed that person.

6.3 IDM and UT behaviour designs

6.3.1 Signalling behaviours

Beyond emotional/status states, there were behaviours developed for navigation and safety. As an example regarding navigation, the LEDs were used to communicate a change of direction, in a way that only the LEDs on the side to which the robot is turning would blink, until it stops turning (Figure 6).

Figure 6. Behaviour of robot indicating changing of direction

In areas where there are a lot of people, the robot will signal that it is aware of its surroundings. This is done through a sequence of alternating the robot's gaze direction from left to right. We have also tested behaviours that would react to the sudden appearance of a person in front of the robot, which allows us to react accordingly, i.e. by changing gaze

attention to the newly appeared person, politely slowing down (even more) or stopping, and using the robot's voice to signal awareness, and it's intention to yield the way of right.

To improve people's awareness of the robot, we added a basic 'electronic-vehicle-sounding' audio clip loop while the robot is driving. The volume and pitch are modulated based on the current robot's driving speed.

6.3.2 Loading/offloading behaviours

For safety, since all the storages have electronic locks, it was established that when a storage is open all the LEDs blink orange, so that people are aware of that information and do not leave the robot without closing all the storages properly, and thus protecting its cargo (Figure 7).

Figure 7. Behaviour of robot signalling when the “storage open”

6.3.3 Resolving obstructions

During delivery tasks, we want the robot to quickly, but politely, resolve situations that keep it from performing its task. People or obstacles obstructing the robot's desired path may be a common event preventing the robot from continuing. We have proposed and implemented a behaviour that attempts to communicate to bystanders that the robot wants to continue driving in a certain direction. The sequence goes as follows:

1. First, rotate itself into the desired driving position
2. Then, address the closest person in the robot's (frontal) field of view by gazing at it with "concerned" eyes, and making a "concerned sound". If there is nobody, briefly "look around" by gaze left, then right, instead.

3. Then, shift gaze to the floor in front of the robot, where it would like to continue its path.
4. Then, in unison, make a beeping noise, flash the front LEDs, make a small motion towards the desired direction (around 5cm), and immediately move back again.

Do 4. twice, then repeat from 2.

We repeat this until the path is not blocked anymore, or until a maximum number of repetitions is reached without the navigation being able to continue. In that case, we need to escalate the behaviour to a different strategy (i.e. notifying staff).

Depending on the context, this behaviour may be expanded more. For example, if there is a static (non-person) obstacle blocking the path, there may be people around, but out of view of the robot. In that case, it may be helpful to first rotate towards that person in step 1, and rotate back to the desired driving direction before step 2.

6.3.4 Basic socially-aware navigation

There are social norms on the (acceptable) distances humans keep between themselves, depending on context and relationship between the individuals. We want the robot to respect these distances at all times. However, in some situations, it may be important to be able to re-establish a safe and socially acceptable distance towards bystanders. Not just after a violation of social space by either robot or bystander, but also to recover a safe distance to continue navigating. A first behaviour has been designed and implemented that demonstrates this ability. When activated, the robot will rotate towards the closest person (if not already facing that direction), and, if the social distance is less than 75cm, the robot will back up slightly, uttering a "sorry" noise. Rotation and motion behaviours are accompanied by LED animations that underline the direction of rotation and movement. Of course, for the field use of this behaviour in particular, this behaviour needs to also take into account safety, i.e. to only back up when there is room behind.

6.3.5 Emotions

To show how to confer some identity to the robot, IDM developed some preliminary displays of emotional states - regular, sad, happy, and frustrated - that can be used by the robot to communicate certain states given a certain context (Figure 8). The LEDs on the base and the “head” complement what is being displayed on the screen, changing accordingly. One of the demonstrative behaviours developed was that when the robot encounters an obstacle (e.g. the robot is in a crowded hall and cannot go past or go around), the LEDs will blink red and the eyes will express an “annoyed” expression (Figure 9). In that way, people in the surroundings will notice that something is wrong and help the robot. Using the same logic, to convey the feeling of sadness, the eyes would change to a sad expression and the LEDs

would turn blue, to convey happiness the eyes would change to a happy expression and the LEDs would have a rainbow-like animation, and so on (Figure 8).

Figure 8. Example of preliminary displays of emotional states - regular, sad, happy, and frustrated.

Figure 9. Behaviour of robot when it encounters an obstacle

6.4 Discussion

We have created an inventory of events where the HRI component could provide added value to the robot's operation. For each event, we have drafted potential behaviour strategies. Test implementations of some of these behaviours have been implemented and tested on the IDM Harmony robot.

The achievements of this integration week provided the foundation for the social behaviours of the HRI component, both in terms of technical implementation and design. The outcomes of the ongoing studies will be informative as to shape these behaviour designs further. Furthermore, we will be able to evaluate these behaviours on the physical robot in a hospital context.

For complete integration, another next step will be working towards an end-to-end demo, where these behaviours are incorporated into the overall robot operation, including the autonomous navigation system and the harmony-specific use-cases for hospital delivery.

7. Mitigation Strategies for IDMINDs Harmony Robot After Navigation Errors

7.1 Introduction

When robots, particularly in environments like hospitals, enter public spaces, their ability to navigate diverse social norms becomes crucial. Each user brings unique expectations, and any deviation from these norms is swiftly perceived as a social error. Hospitals, inherently high-stress environments, continually subject patients and healthcare professionals to situations that trigger anxiety, depression, fatigue, and pain. While service robots promise enhanced efficiency, their missteps can create confusion and, more critically, pose safety risks. An example highlighting this issue is evident in studies where unexpected behaviour from a hospital delivery robot perplexed nurses, leading to a blame game among them, the robot, and their peers for the robot's conduct (Kim et al., 2006). This underscores the importance of robots not only acknowledging their mistakes but promptly addressing them to prevent escalating confusion.

The realm of human-robot interaction has explored strategies for rectifying social errors, including apologies, explanations, promises, and compensations (Esterwood and Robert, 2022). However, a challenge we are grappling with regarding the Harmony robot lies in its non-verbal communication approach, as mentioned in previous sections. While this lowers user expectations to a more realistic standard, it also complicates conveying complex messages. Hence, we explore non-verbal strategies for communicating acknowledgment of a mistake, such as expressing regret or a promise of having learned from the error, as the robot's error mitigation recovery within a hospital setting for three different errors that occur during navigation.

Prior work has found that the effectiveness of a mitigation strategy changes depending on the perceived severity of the error (Stiber and Huang, 2020) and people’s expectations of the robot’s functionality (Washburn, 2020). Because people’s expectations of the functionality of the Harmony robot can change depending on the role they have in the hospital (i.e., patient, visitor, staff), we are interested in researching if the effectiveness of the mitigation strategies changes across these roles. Additionally, previous research found that exposure to errors multiple times can affect the effectiveness of mitigation strategies (Esterwood and Robert Jr., 2023). Because staff will interact with the robot more than visitors and patients, we also want to study how the mitigation strategies (promise vs. regret) are effective over time (when multiple errors occur). This study aims to examine the development of behaviours to mitigate social navigation errors and add these to the Harmony robot’s dictionary, which the robot will later use to decide which behaviour to employ given the contextual requirements. An online study will be conducted to evaluate

multimodal recovery strategies across multiple errors and users with different roles within a hospital environment.

7.2 Design of Video Vignettes

This study involves shooting video vignettes within a virtual reality (VR) environment. These vignettes feature a simulation of the IDMinds Harmony robot enacting three social errors during navigation, two error recovery strategies, and one control condition where there is no error and no recovery strategy.

7.2.1 Display of Social Capabilities

For participants to have a baseline on IDM Harmony robot's behaviour during navigation, we will record a video from a third-person perspective where the robot navigates a hospital corridor and encounters different social moments with virtual humans.

7.2.2 Social Errors

The vignettes will be recorded from a third-person perspective to increase the possibility of the participants observing the errors clearly, as errors that happen close to them could be hard to notice because of their perspective (the robot could be too close to the camera for them to notice nuances between some of the mistakes). These errors, nine in total, will be specifically designed for the robot's actions in a hospital corridor. For instance, one error involves the robot inadvertently encroaching on someone's personal space, another interrupts conversations between individuals in the corridor, while a different scenario depicts confusion when the robot and a person walking towards each other block the corridor, causing a standoff.

7.2.3 Recovery Strategies

The videos of the recovery strategies will be recorded in first person perspective. This is to ensure that the participants can see the eye expressions and movements the robot uses. To communicate regret, we looked into prior work in both human-human and human-robot interaction. Here we sound that showing sadness and embarrassment can help convey this emotion. To recover from a social mistake by employing humour, we focused on designing joyful unexpected behaviours for the Harmony robot.

Because we are interested in seeing how the recovery strategies are perceived with the use of different modalities (minimal vs. complex vs. with the use of sound), we designed three ways to convey the recovery strategies (humour and regret) using these types of use of modalities. Hence, we end up with six different videos, three per recovery strategy, one with

the use of minimal modalities, one with complex multimodal behaviours, and one using sound.

In addition to the two behaviours to convey the mitigation strategies, we also designed a control condition where the Harmony robot does not detect that an error happens and just continues its path down the corridor.

7.3 Pre-study

To identify three videos that diminish participants' perception of the Harmony robot's social intelligence following an error, we devised a preliminary study. In this pre-study, participants will assess the robot's perceived social intelligence and the severity of each mistake made.

7.3.1 Method

7.3.1.1 Participants

We expect to recruit around 30 participants through convenience sampling. Participants will only be eligible to participate in the study once; they must be older than 18 years old, understand English, and have access to an electronic device to take the survey, along with speakers/headphones. We will ask participants to provide us with demographic information such as their gender identity, nationality, previous experience with robots, and age.

7.3.1.2 Procedure

Participants will initially use an online questionnaire to rate the Harmony robot as it demonstrates its social capabilities while interacting with users in a hospital corridor. Following the viewing of this video, participants will be prompted to rate the robot's perceived social intelligence.

Then, participants will assess nine different videos depicting the Harmony robot committing a social error in a hospital corridor. The sequence of these error videos will be randomised. After each video, participants will rate the robot based on its perceived social intelligence and the severity of the mistake the robot made. As a manipulation check, we will also ask participants to describe the error they saw during the video as an open question. Lastly, they will answer demographic questions.

7.4 Main Study

The goal of the main study is to assess how participants rate IDM Harmony robot’s social intelligence, how it could change after it makes a social mistake, and how it could change again after it uses one of our designed recovery strategies.

7.4.1 Method

This study contains three different variables: social error, recovery strategy, and design of the recovery strategy. All participants will experience the three recovery strategies, and the three different designs for them (as they are within-participant variables). However, they will not get to experience all the combinations of errors and recovery strategies (as these will be between participants).

7.4.1.1 Participants

For the main study, we will use Prolific to recruit our participants. We aim to recruit 400 participants. Participants will only be eligible to participate in the study once; they must be older than 18 years old, understand English, and have access to an electronic device to take the survey, along with speakers/headphones. We will ask participants to provide us with demographic information such as their gender identity, nationality, previous experience with robots, and age. The survey is expected to take 30 minutes and participants will be rewarded $9 U.S. dollars for their participation.

7.4.1.2 Stimuli

During this study we will use the robot social errors selected from the pre-study, as well as the errors in combination with a mitigation strategy or no mitigation strategy.

7.4.1.3 Measures

To assess the robot's perceived social capabilities before, during, and after an error, as well as post-recovery, we'll employ the Social Competence (SoC) (Bachard, 2020) subscale within the Perceived Social Intelligence (PSI) scale. This scale comprises four items: This robot... 1) is socially competent, 2) is socially aware, 3) is socially clueless, and 4) has strong social skills.

To assess the legibility and appropriateness of the recovery strategies, participants will rate the robot's behaviour in terms of how sorry, funny, and appropriate it was following a mistake, using a 7-point Likert scale.

As a means of checking how participants perceived the error depicted in the video, we'll prompt them to describe the specific error made by the robot in an open-ended question.

7.4.1.4 Procedure

The survey will be online and participants will be able to access it through Prolific. It will show participants an information letter about the study and its goals and ask for their

consent to participate.

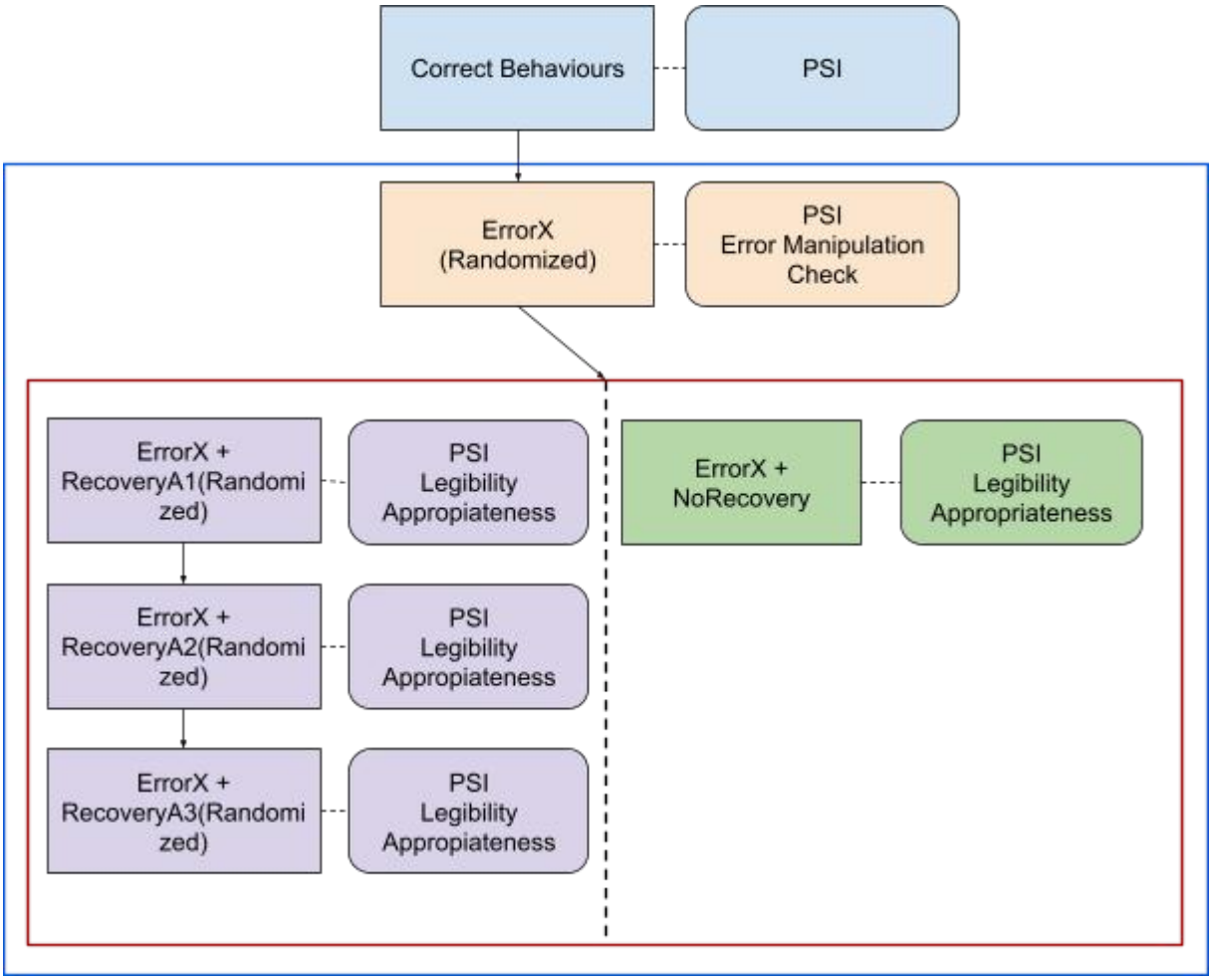

Figure 1. Diagram of the study procedure with measures.

To start, participants will view a video featuring the well-behaved robot and rate its PSI. Following this, participants will be informed that they will watch a video where a robot makes an error. They'll view one error scenario (randomised and balanced) and then proceed to rate the robot's PSI. Additionally, they'll answer an open question regarding the observed error. Subsequently, participants will be notified about viewing the same error three more times, with interspersed questions between each viewing. During these repetitions, participants will witness the error vignette accompanied by one of the two recovery strategies or no mitigation (randomised and balanced). In instances where an error is paired with a recovery strategy, the strategy will be presented in one of the three different designs we've developed: minimal, complex, or sound (randomised and balanced). After each video viewing, participants will rate the robot's PSI, as well as provide legibility and appropriateness ratings. Following this, they'll watch the same error with a different (randomised and balanced) recovery strategy design and respond to similar questions, except in the case of no recovery. Once participants experience the three recovery strategy designs for each error scenario and rate the PSI, legibility, and appropriateness for each

video, or in the case of no recovery, they'll be presented with a new error. They'll rate only the PSI and respond to open-ended questions for the error manipulation check. This sequence will continue until each error is shown once. In total, participants will watch seven videos, encompassing three errors, two recovery strategies, and one no-recovery strategy video.

Lastly, after participants go over the seven videos, they will be prompted to answer some demographic questions.

7.5 Discussion

This study will be run during the winter of 2024 (January and February), we plan to recruit online participants through the Prolific recruitment tool. In addition to this online evaluation, we also plan to run a more qualitative version of this study at UZH with hospital staff. Results for both of these studies will be reported and discussed in Deliverable 8.5.

Appendix

Appendix A - Event List

Event	Sub event	Resolution	Exit condition
Obstruction	Object blocking the way, no people in proximity	Signal need for help	Obstruction resolved; direct to person; signal "thanks"; continue path
	Object(s) blocking the way, person in proximity	Turn to person, signal need for help, point towards obstruction	Obstruction resolved; direct to person; signal "thanks"; continue path
	Object(s) blocking the way, people in proximity	Turn to random person, signal need for help, point towards obstruction	Obstruction resolved; direct to previous random person; signal "thanks"; continue path
	"Other/Undefined"	Signal need for help	Obstruction resolved; direct to person; signal "thanks"; continue path
Elevator	entering, people in elevator	Signal hello	Directly after resolution
	leaving, people in elevator	Signal bye	Directly after resolution
	Stop on a floor that is not the destination floor, person leaves	Signal bye	Directly after resolution
	people entering elevator when robot is inside	Signal hello	Directly after resolution
People interrupting trip, but no obstruction	People interrupting trip, but no obstruction	Signal sorry and signal urgency, then move on with task	Directly after resolution
Person needs the robot to move out of the way	see "obstruction"	-	-
Navigation encounters	Passing near a corner area	Signalling its presence	Directly after resolution
	Driving close to people	Slow down (perhaps TUD's job to integrate in their module)	Directly after resolution
	Driving Alone	Speed up	Directly after resolution
	Corner turn, person surprise	Mitigate mistake	Person moves on/or timer completed
	People with mobility aids	Slow down and acknowledge person	Directly after resolution
Outside of the map (lost)	Outside of the map (lost)	Signal need for help + more complex signalling to explain robot is lost?	Until helped. Repeat if it takes too long. Exit when a map is found and planning is made.
Robot gets stuck on the floor	Robot gets stuck on the floor	Signal need for help	Until helped
Harmony specific	Robot is ready to be loaded	Signal "hello" and "ready to be packed"	Repeat every now and then until person interact with robot
	Robot is fully loaded and can start delivery	Signal "thanks" and "bye"	Directly after resolution
	Robot arrived at destination and can be unpacked	Signal "hello" and "ready to be unpacked"	Repeat every now and then until person interact with robot
	Robot unpacked and returning to delivery point	Signal "happy" and "bye", then start moving	Directly after resolution

References

Barchard, K. A., Lapping-Carr, K., Westfall, R. E., Fink-Armold, A., Balajee Banisetty, S., & Feil-Seifer, D. (2020). Measuring the Perceived Social Intelligence of Robots. J. *Hum.-Robot Interact*. 9, 4, Article 24 (December 2020), 29 pages. https://doi.org/10.1145/3415139

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F. & Bull,S. (2013). Towards Empathic Virtual and Robotic Tutors. 7926. 733-736. 10.1007/978-3-642-39112-5_100.

Dunne, A., & Raby, F. (2013). *Speculative Everything: Design, Fiction, and Social Dreaming*. MIT Press.

Esterwood, C. & Robert, L.P. (2022) A Literature Review of Trust Repair in HRI. In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE Press, 1641–1646. https://doi.org/10.1109/RO-MAN53752.2022.9900667

Esterwood, C., & Robert Jr, L. P. (2023). Three strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior*, *142*, 107658. https://doi.org/10.1016/j.chb.2023.107658

Goffman, E. (1967). Interaction ritual: essays on face-to-face interaction. *Aldine*.

Groff, E. (1995) Laban Movement Analysis: Charting the Ineffable Domain of human Movement, *Journal of Physical Education, Recreation & Dance*, 66:2, 27-30, DOI: 10.1080/07303084.1995.10607038

Jee, E.-S., Kim, C. H., Park, S.-Y., & Lee, K.-W. (2007). Composition of Musical Sound Expressing an Emotion of Robot Based on Musical Factors. *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 637–641. https://doi.org/10.1109/ROMAN.2007.4415161

Kim, L. H., & Follmer, S. (2021). Generating Legible and Glanceable Swarm Robot Motion through Trajectory, Collective Behavior, and Pre-attentive Processing Features. *ACM Transactions on Human-Robot Interaction*, *10*(3), 1–25. https://doi.org/10.1145/3442681

T. Kim & P. Hinds (2006). Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction. *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*.

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots, *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Christchurch, New Zealand, 2016, pp. 463-464, doi: 10.1109/HRI.2016.7451807.

Read, R., & Belpaeme, T. (2014). Situational context directs how people affectively interpret robotic non-linguistic utterances. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 41–48. https://doi.org/10.1145/2559636.2559680

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A Questionnaire of Cognitive and Affective Empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Savery, R., Rose, R., & Weinberg, G. (2019). Finding Shimi's voice: fostering human-robot communication with music and a NVIDIA Jetson TX2. *Proceedings of the Linux Audio Conference 2019*, 101–105. https://lac.linuxaudio.org/2019/

Schönbach, P. (1980). A category system for account phases. *European Journal of Social Psychology*, 10(2), 195–200. https://doi.org/10.1002/ejsp.2420100206

Stiber, M., & Huang, C.-M. (2020). Not all errors are created equal: Exploring human responses to robot errors with varying severity. *Companion Publication of the 2020 International Conference on Multimodal Interaction*. https://doi.org/10.1145/3395035.3425245

Tian, L., & Oviatt, S. (2021). A taxonomy of social errors in human-robot interaction. *ACM Transactions of Human-Robot Interaction, 10*(2), Article 13. https://doi.org/10.1145/3439720

Washburn, A., Adeleye, A., An, T., & Riek, L. D. (2020). Robot errors in proximate HRI. *ACM Transactions on Human-Robot Interaction*, *9*(3), 1–21. https://doi.org/10.1145/3380783

Yilmazyildiz, S., Henderickx, D., Vanderborght, B., Verhelst, W., Soetens, E., Lefeber, D. (2011). EMOGIB: Emotional Gibberish Speech Database for Affective Human-Robot Interaction. In: D'Mello, S., Graesser, A., Schuller, B., Martin, JC. (eds) *Affective Computing and Intelligent Interaction*. ACII 2011. Lecture Notes in Computer Science, vol 6975. Springer, Berlin, Heidelberg

Yilmazyildiz, S., Read, R., Belpeame, T., & Verhelst, W. (2016). Review of Semantic-Free Utterances in Social Human–Robot Interaction. *International Journal of Human-Computer Interaction*, *32*(1), 63–85. https://doi.org/10.1080/10447318.2015.1093856

Zaga, C., de Vries, R. A. J., Li, J. J., Truong, K. P., & Evers, V. (2017). A Simple Nod of the Head. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, *May*, 336–341. https://doi.org/10.1145/3025453.3025995