Remote-HRI: A Pilot Study to Evaluate a Methodology for Performing HRI Research During the COVID-19 Pandemic

Curtis L. Gittens
The University of the West Indies Cave Hill Campus
P.O. Box 64, Bridgetown, Barbados
curtis.gittens@cavehill.uwi.edu

Abstract

The essential component of studying how humans and robots relate is to observe how they interact with each other. This fact has been taken for granted until the COVID-19 pandemic made all such HRI study-based interactions verboten. We address the question of how HRI research can continue in an environment where physical distancing is the most effective approach to containing the spread of the disease. In this paper we present a pilot study that indicates that Remote-HRI (R-HRI) studies can be a feasible alternative. The results suggest, importantly, that whether the participant interacts with the robot in-person (but physically-distanced) or remotely, their perception of the robots is unaffected. We observed increased effort and frustration by participants when interacting remotely with the robot – for which there may be an underlying cultural cause. However, all participants completed their tasks with the robot and had a neutral to positive experience using it.

1. Introduction

With a global pandemic affecting all aspects of society, research into human-robot interactions have faced the problem of how to undertake studies while maintaining public health requirements. Although countries in Europe and Asia have reopened, there are still spikes in infection rates occurring in some areas, and governments remain on alert for possible future outbreaks. In the USA, many states remain open, but with uncontrolled community spread, government officials are beginning to roll back their previous reopening plans.

The SARS-Cov-2 virus that causes COVID-19 is highly contagious and very dangerous to the elderly and those with underlying conditions. Unfortunately, these populations are the ones in which most HRI research is focused – as a simple Google Scholar search would attest. Recognizing this problem, Feil-Seifer et al. [1] identify nine research questions that must be answered so that human-oriented research such as HRI can continue within the context of this pandemic.

One area they identify is the need to develop methodologies that would facilitate HRI research in a practical and ethical way during the pandemic [1]. This is crucial since significant work on assistive and companion robots is done with the elderly and populations with underlying health conditions [2]–[8] and it is these populations that are most at risk if they are exposed to the virus [9].

1.1. Remote Usability Testing

We believe that the answer to the question posed by Feil-Seifer et al. can be addressed by the work done in Remote Usability Testing (RUT) [10], [11].

Conventional usability testing involves gathering participants, which represent the target user group of the software under test, into a lab where there is a moderator to observe them as they work on specified tasks to identify usability errors [12]. The protocol for usability tests involves the participant being encouraged to talk aloud as they work through their task. This permits the moderator to gain some insight into the participant's thought processes, and to identify any misunderstandings that may occur as the user interacts with the system [12].

While this method has a number of advantages – a common one being high-quality quantitative data – significant drawbacks also occur. Impediments include, high costs, difficulty in creating environments that represent end users' environment, and most importantly, identifying a representative sample of participants that can physically come to the lab. The last problem is exacerbated by the fact that software is sold internationally. This means that a representative sample of participants must be drawn from across the globe wherever the software is sold [13].

RUT addresses these problems. Video conferencing tools like Zoom®, MS Teams® and Google Meet® provide the real-time interaction needed. Each of these applications have screen sharing and other collaborative tools to facilitate interaction with the moderator and the participant. Prior to these systems, software such as Cisco WebEx® and GoToMeeting® provided similar functionality. Both unmoderated and moderated remote usability testing approaches are used, and studies have shown the benefits of both [14]–[16].

The lessons learned from, and applied to, RUT provide a roadmap for undertaking remote human-robot

H┇CSS

interaction experiments in a pandemic. Firstly, studies have shown that RUT is as effective as the conventional in-lab approach [14], [16]. More recently it has been shown that the lab environment can be replicated as a virtual environment without any significant workload differences from the conventional lab or even a remote testing approach [17].

1.2. Remote Human-Robot Interaction

While there are remote-controlled robots, this is not what is being proposed in this work since a remotely-controllable robot still needs to be tested before it can be deployed. In this paper, we propose a Remote Human-Robot Interaction (R-HRI) study methodology where experiments are either done where: (a) the participant is present but physically distanced from the robot or; (b) the participant is physically remote.

In the first case, all public health requirements pertaining to controlling the transmission of viruses and specifically coronaviruses are enforced [18], and includes not touching the robot. This is especially important for team-based studies or studies where different participants are usually required to touch the robot. Since disinfecting the robot could potentially damage its screen or other interfaces, the moderator acts as the intermediary for the participant and the robot in these circumstances. The key difference between this role and that of a RUT moderator is, whereas the usability testing moderator encourages elicitation of thought processes, the R-HRI moderator's role is more like that of a proxy for the participant. That is, they help the user perform tactile or other interactions that would violate public health protocols and do so without attempting to elicit information on the user's thought processes.

In the fully-remote case, the participant and the robot are in different locations. Being able to perform experiments under such conditions is important when HRI studies involve at-risk populations, where going into their living spaces is prohibited. In this configuration, the moderator plays a similar role to that of a moderator in RUT, with the modifications to the role as described above.

We designed an experiment to determine whether R-HRI is feasible. This feasibility is evaluated from the human attitude, cognitive workload and human experience perspectives and focus on three questions:
1. Will a human participating in an R-HRI experiment manifest the same attitudes towards robots regardless of whether they interact with the robot in-person or remotely?
2. Will the cognitive workload of getting the robot to perform an action be the same whether interacting with the robot remotely or in-person?

3. Will the user's experience with the robot be the same whether it is in-person or remotely? If not, are the aspects of the experience in an R-HRI study so different that it would be infeasible to use?

We undertook a pilot study to provide initial answers to these questions. We did a pilot study as a precursor to a full study for a number of reasons. Firstly, we had to determine what potential impact the supporting technologies, network, cameras, speakers, etc., will have on the R-HRI experiment. While we did not focus on technology design in this study, we informally made note of its configuration and potential effects. Additional details on this are provided in the Discussion section. Secondly, we wanted to determine if these forms of R-HRI were indisputably negative experiences. Undertaking a full study without first determining the merit of the idea would be unwise.

2. Related Work

The term Remote-HRI was used by [19] to describe how participants in the study by Huber and Weiss used a remote-controlled, off-the-shelf robotic arm. In their study, the term was applied to operating the robot with a remote-control, and this was evaluated against Physical-HRI – the term they used to describe participants manually controlling the robot. The Huber-Weiss definition differs from the definition in the work presented here because our definition focuses on the nature of the physical relationship the user has with the robot during the experiment - specifically that of the separation – and not as a method of controlling the robot.

Most research pertaining to remote robot interaction focuses either on how to use a robot to interact with a remote environment [20], [21] or how to manipulate the robot from a remote location (teleoperation) [22], [23]. There have been some studies that focus on the success of HRI outcomes by establishing common ground between human and robot. Work done by Stubbs et al. [24] discusses the use of a robot proxy – a system that simulates the robot's responses – to improve interaction outcomes by establishing common ground between the human and robot. While HRI was part of the study, the main focus of this work was on the effectiveness of the grounding process between the human and robot. As a result, issues regarding physical separation, attitudes and experience are not examined. The case is similar for other work on robot proxies and grounding [25], [26].

Most recently, work done by Honig and Oran-Gilad investigated interaction using online video surveys to evaluate the quality of hand gestures used to communicate with robots [27]. The study was motivated by some of the same factors that motivate RUT, namely

the cost and effort associated with mounting in-person studies – especially in situations where the robot has not been fully tested for interaction with humans. The study did not use a moderator, nor was real-time interaction used; however, the researchers had participants watch videos and respond using a video – asynchronously. The results showed that the method of interaction had no effect for 7 of the 8 selected gestures and that their approach might a viable alternative to in-person testing especially during the prototyping stage. Their assessment is similar to the investigation presented in this paper.

The area of human-robot proxemics investigates physical- and psychological distancing between humans and robots [28], [29]. The focus of this area of research is in determining how humans react to robot's behaviors when they are within a certain distance from the user, as well as determining how a robot can affect human behavior or determine human emotions and other types of affect. Human-robot proxemics does not directly address the questions we seek to answer in this paper. While physical distancing is studied in proxemics, this is done within the context of making the robot model social and cultural norms. Physical distancing in our work is a constraint under which experiments must be executed. So, we do not investigate physical distance by itself, as we would with a human-robot proxemic study [28], but instead, preset a physical distance from the robot and perform HRI studies with that condition.

Although there is previous work that focuses on the outcomes of the nature of interaction the user has with a robot, those works have not investigated how these interactions function in a physically-restricted setting. They also have not investigated the impact these interactions have on user attitudes, workload and experience. In this paper, we address these restrictions and their impact on interaction by undertaking a pilot study to determine the viability of Remote-HRI as a way to undertake HRI studies within the context of a global pandemic where physical interaction is limited.

3. The Method

3.1. Participants

The pilot study was done with participants consisting of seven males between the ages of 22 and 46 ($\mu = 32.3, \sigma = 7.36$) and one moderator (male, 22). The participants were recruited from a computer science research group, whose primary areas of research do not involve social robots. Before their participation in this pilot study, four out of the seven participants (57%) had prior interaction with some robot that was different than the one used in this study. We used the advice given by Macefield [30] where a baseline of 5-10 participants

being a sensible baseline for early usability studies. We believed this advice could apply to our study.

3.2. The Experiment Conditions

Mode of interaction with the robot was the independent variable. It was evaluated at two levels: in-person interaction with a moderator, and remote interaction with a moderator.

The in-person environment was designed to meet all public health requirements that are known to reduce the likelihood of human-to-human transmission of infections [18]. Participants were required to wear face masks and stand no less than two meters away from the robot and moderator. Any interaction with the robot that required touch had to be done through the moderator, who was the only person allowed to touch the robot. The in-person tests were done in an open outdoor courtyard to reduce the possibility of aerosolized particles circulating in the test location – even though airborne transmission of the virus has not been verified [18]. This environment also facilitated remote interaction with the robot. A laptop with a built-in webcam and augmented with an external Bluetooth speaker was used to run the video conferencing application. The robot was placed in front of this system so that the remote user can see Zenbo move around and so that the robot can hear the remote user's commands.

We used a computer research lab as an on-site remote environment. The lab accommodates six researchers; however, no more than three people were allowed to be in the lab at any given time to facilitate adequate spacing. Gloves and hand sanitizer were provided to participants – if they chose to use the laptop provided to interact with the robot remotely. Face masks were required while inside of the lab.

3.3. The Robot

The ASUS Zenbo companion robot was used in this study, see Figure 1. The robot is a multi-purpose intelligent assistant that has been described as cute [31]. It is designed to work in educational, business and health and safety environments with roles such as: a teacher assistant's storyteller; a companion robot for the elderly and; a retail service assistant [32]. Given the intended uses for the robot, we believed they made it a good candidate for operation by first-time users.

We chose to use some of the Zenbo robot's built-in functions since these applications have been tested by other researchers [33]. We chose these applications to reduce confounding that may occur with custom applications that are less established, less tested and could introduce unexpected outcomes in this pilot study.

Figure 1. The Zenbo intelligent robot.

For each experiment, we had each participant go through a pre-selected list of nine functions available with Zenbo. These functions are listed in Table 1. We chose these functions because the robot's expected reactions were easy to describe to participants. The chosen commands are also simple and we anticipated that the commands that there would be easy for participants to issue.

We also required a range of commands to have users do testing with Zenbo. The robot offers oral and movement responses. The user can also interact with

Table 1. List of Zenbo commands numbered 1-9

Command	Function	Expected Reaction
1) "Hey, Zenbo"	Prepare to receive a command	Blue "ears" appear and waits for a command
2) What can you do?	Opens list of functions	Display functions
3) "What date is it tomorrow" OR "What date is it tomorrow on the lunar calendar"	Date report	Tomorrow's date in the calendar/lunar calendar
4) "What is the weather in Barbados today?" OR "What is the weather today?"	Weather report	Reports the weather
5) Stroke its head (ask moderator)	Basic interaction	Shows a shy expression
6) "Follow me"	Following	Follows the user
7) "Tell me a story"	Entertainment	Tells the user a story
8) "I want to take a picture" / "I want to take a selfie"	Photo	Takes a picture of the user
9) "I want to listen to music	Entertainment	Plays music

Zenbo using voice directives or touch. Of the nine instructions, five commands (1, 2, 3, 4 and 9) were voice directives that resulted in an oral response from Zenbo. Three commands (6, 7, 8) were voice directives that resulted in an oral and movement response by Zenbo. One command (5), required physically touching the robot and resulted in movement by the Zenbo robot. For all commands except Command 5, the participants interacted with Zenbo themselves. However, to perform Command 5, the participant could not touch the robot but instead had to ask the moderator to perform the action.

3.4. Devices Used for Remote Access

Participants had three options to access the robot remotely: (i) use the laptop provided; (ii) use their personal laptop or; (iii) use their smartphone. The laptop provided to participants ran the Chrome web browser and the remote desktop application AnyDesk®. We used the Jitsi [34] video conferencing website to allow participants to communicate with the Zenbo companion robot, see it move around and to communicate with the moderator. The AnyDesk application was used to give the participants a clear picture of the robot's face and its expressions because these were hard to discern through the video conferencing website. AnyDesk was installed on Zenbo to make this type of connection possible. Participants who chose to use their own laptop had to install AnyDesk and then use the application to request a connection to the robot. The moderator ensured this was done at the start of the session. The final device option was for participants to use their smartphones. With this option participants had to install the Jitsi application to get the best quality connection. Additionally, using the smartphone prevented participants from using the AnyDesk application since it was not possible to use both the Jitsi and AnyDesk applications simultaneously. Using this arrangement, participants saw the robot's expressions only through the Jitsi video conferencing application.

3.5. The Survey Instruments

For this study we selected five survey instruments that could help us determine whether: (a) participants' mode of interaction with the robot (in-person/remotely) affected their attitudes towards the robot; (b) the effort expended in interacting with the robot depended on interaction modality and; (c) the quality of the participants' experience with the robot was affected by the interaction mode.

We used the Negative Attitude towards Robots Scale (NARS) [35] to determine the level of negative perceptions held by the participants towards robots.

NARS is a 14-item scale that has three subscales. The first, six-item subscale, measures negative attitudes towards human-robot social interactions (HRSI). The second, five-item subscale, measures the negative attitudes towards the social influence of robots (SIR) and the third, three-item subscale, measures negative attitudes towards emotional interactions with robots (EIR). Measuring negative perceptions was important since there are very few companion, assistive or service robots commonly available in our region. This meant that any impression participants may have had about robots would likely not have been from prior exposure, but from second-hand sources like popular media. The higher the NARS score, the more negative the participant's attitude towards robots. This means that by using this scale we could determine whether participants' attitudes change after each interaction with the robot (Pretest: α = .76; Posttest I: α = .66; Posttest II: α = .68).

The Robotic Social Attributes Scale (RoSAS) [36] was used to get a better sense of the perceptions held by participants, especially if NARS showed that participants did not have any negative attitudes towards robots. RoSAS is an 18-item scale with three subscales: *Warmth*, *Competence* and *Discomfort*, each consisting of six items. The higher the score in each subscale, the greater the perception by the user of the robot possessing the characteristics described by the items of the subscale. For example, a robot that is perceived as warm, implies that the participant scored it highly in the six items: happy, feeling, social, organic, compassionate and emotional that comprises the warmth subscale. We used this scale to determine if any of these perceptions changed based on the interaction modality used with the robot (Pretest: α = .89; Posttest I: α = .78; Posttest II: α = .75).

In addition to human perception, we also assessed use of the robot in an effort to capture whether the interaction affected levels of technology adoption. We used the Extended Technology Acceptance Model (TAM2) [37] scale to determine participants' attitudes towards accepting a companion robot for use if one were available. TAM2 measures technology acceptance, and we regarded it as a viable indicator of the quality of the interaction with the robot. We assessed the factors: Intention to Use (IU); Perceived Usefulness (PU) and Perceived Ease of Use (PEU). The higher the TAM2 score, the more likely the participant is to accept and/or adopt the technology under assessment (In-Person: α = .82; Remotely: α = .82).

The NASA Task Load Index (TLX) consists of six subscales: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E) and Frustration (F) [38]. Its total score measures the workload associated with a specific task. A key measure

we wanted to capture was the quality of the participants' experience with the Zenbo robot within the constraints of the experiment environment. To achieve this, we examined the *Performance*(P), *Effort*(E) and *Frustration*(F) factors of the TLX with the two modes of interaction (In-Person: α = .81; Remotely: α =.74).

The short version of the User Experience Questionnaire [39] is a seven-point, eight-item inventory that allows subjects to provide a full assessment of their experience using a technology. It has two subscales: *Pragmatic Quality*, which measures how efficiently you can perform the task using the product; and *Hedonic Quality*, which measures how interesting and stimulating it is to use the product to perform the task. Values greater than 0.8 represent a *positive* evaluation, values less than -0.8 represent a *negative* evaluation. Values between -0.8 and 0.8 represent a neutral evaluation. Given this, we can determine the quality of the participant's experience after each interaction mode (In-Person Pragmatic: α = .95; Remotely Pragmatic: α = .87); (In-Person Hedonic: α = .97; Remotely Hedonic: α = .85).

3.6. The Experiment Design

We used a counterbalanced measures design with the independent variable (interaction mode) at two levels: in-person and remote. The seven participants were randomly assigned to either Group A (which first interacted with the robot in-person and then remotely) or Group B (which interacted with the robot remotely first, and then, in-person). Data was collected over a period of nine hours.

3.7. Procedure

Before starting the experiment, participants from the research group were given an online copy of the informed consent form, which they had to read and then click to accept. After giving consent, basic demographic information such as age and whether that had ever interacted with a robot was collected. Participants then filled out the NARS and RoSAS inventories, and were instructed on how to issue the nine commands shown in Table 1 to interact with the robot. After their first interaction, they were instructed to fill out the five inventories. When they completed the surveys, they used the same commands to interact with the robot a second time using a different interaction modality. For example, a participant that was assigned to Group A, would complete the NARS and RoSAS inventories in the pretest, interact with the Zenbo robot in-person, and then complete the five inventories in posttest. After their second interaction, the five inventories are again completed in the second posttest. The process is the

same for Group B participants who start with the remote interaction mode.

4. Results

Given the small sample size of the pilot study, we used the Friedman's non-parametric test on the NARS and RoSAS data for within subject analysis of the effects of the interaction modalities against attitudes towards robots. To analyze the other scales that measured technology adoption (TAM2), workload effort (NASA TLX), and user experience (UEQ), we tested for normality using a visual test (Q-Q Plot) verified by a Shapiro-Wilk test, which has more power than the Kolmogorov-Smirnov normality test [40]. If the data met the normality requirement, we investigated the presence of effects using a repeated measures *t*-test, since it functions with small sample sizes [41]. Data samples that failed the normality tests were analyzed using the Wilcoxon Signed-Rank test.

4.1. NARS

We performed the Friedman's Test on the three subscales of the NARS separately. The three treatments were the pre-interaction versus in-person interaction versus remote interaction against the independent variable *interaction modality* at two levels: in-person and remote. For the HRSI subscale there was no significant effect of the interaction modality on user attitudes, $X^2(2, N = 7) = 3.43, p = 0.18$. For the SIR subscale, there was no significant effect of the interaction modality on social interaction attitudes, $X^2(2, N = 7) = 0.5, p = 0.78$. Finally, for the EIR subscale, there was no significant effect of the interaction modality on attitudes towards emotional interactions with robots, $X^2(2, N = 7) = 0.28, p = 0.87$.

4.2. RoSAS

We performed a Friedman's test on the three subscales of the RoSAS inventory. The three treatments were the pre-interaction versus the in-person interaction versus the remote interaction against the interaction modality at two levels: in-person and remotely. For the *Warmth* subscale there was no significant effect of the interaction modes on this attitude, $X^2 = (2, N = 7) = 0.5, p = 0.78$. For the *Competent* subscale there was no significant effect of the interaction modalities on the attitude, $X^2 = (2, N = 7) = 2, p = 0.37$. For the *Discomfort* subscale, there was no significant effect of the interaction modalities on this attitude, $X^2(2, N = 7) = 1.36, p = 0.51$.

(a)

(b)

(c)
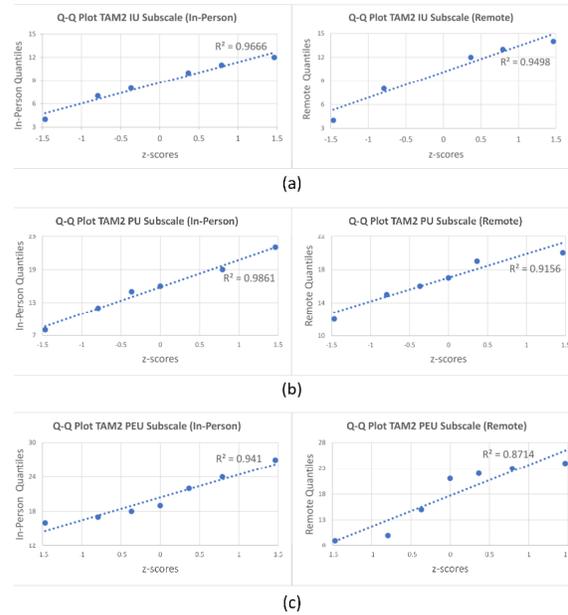
Figure 2: Q-Q plots of (a) TAM2 IU subscale; (b) TAM2 PU subscale and; (c) TAM2 PEU subscale showing that the subscale data may fit a normal distribution, with minimum R^2 value of 0.871, 2(c), and maximum R^2 value of 0.9861, 2(b).

Table 2: Statistics and Normality Tests for TAM2

	In-Person (N=7)	Remotely (N=7)
Subscale: Intention to Use (IU)		
Mean	8.571	9.571
S.D.	2.7	3.5
Shapiro-Wilk Normality Test		
W	0.960	0.912
p-value	0.92	0.46
Subscale: Perceived Usefulness (PU)		
Mean	15.571	16.857
S.D.	4.791	2.794
Shapiro-Wilk Normality Test		
W	0.979	0.935
p-value	0.995	0.683
Subscale: Perceived Ease of Use (PEU)		
Mean	20.429	17.714
S.D.	4.035	6.317
Shapiro-Wilk Normality Test		
W	0.931	0.850
p-value	0.643	0.139

4.3. TAM2

For the TAM2 scale, we used Q-Q plots to perform normality visual tests. They appeared to indicate that the data for all three subscales were normally distributed, Figure 2. We ran the Shapiro-Wilk normality for

additional verification, Table 2, which allowed us to use the *t*-test for all subscale analysis. We performed a repeated measures *t*-tests on the *Intention to Use* (IU) subscale from the TAM2 inventory. The results indicated that the interaction modality had no effect on Intention to Use $t(6) = .733$, $p = .49$.

We performed a repeated measures *t*-test on the *Perceived Usefulness* (PU) subscale and the results indicated that the interaction modality had no effect on Perceived Usefulness $t(6) = .668$, $p = .53$. Finally, we performed a repeated measures *t*-test on the *Perceived Ease of Use* (PEU) subscale and the results indicated that the interaction modality had no effect on Perceived Ease of Use $t(6) = -1.056$, $p = . 33$.

4.4. NASA Task Load Index

For the NASA TLX scale, we used Q-Q plots to perform normality visual tests that appeared to indicate that the data for all three subscales were normally distributed, Figure 3. We ran the Shapiro-Wilk as a verification test for normality, Table 3, which allowed us to use the *t*-test to analyze the *Performance* and *Frustration* subscales. The *t*-test normality requirement
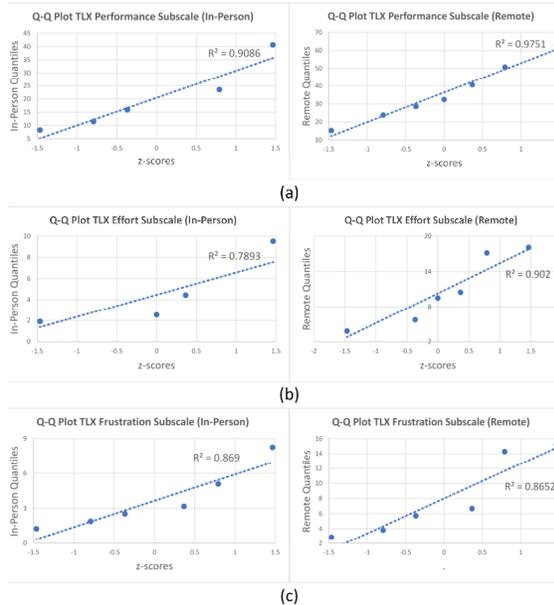
Figure 3: Q-Q Plot of NASA TLX (a) *Performance* subscale; (b) *Effort* subscale and; (c) *Frustration* subscale showing that the data may fit a normal distribution.

was not met for the In-Person interaction modality for the Effort subscale, so we used the Wilcoxon Signed-Rank test.

The result of the repeated measures *t*-test for the interaction modality showed that it had no effect on

Table 3: Statistics and Normality Tests for NASA TLX

	In-Person (*N*=7)	Remotely (*N*=7)
Subscale: Performance		
Mean	18.775	36.372
S.D.	10.724	16.461
Shapiro-Wilk Normality Test		
W	0.824	0.979
p-value	0.081	0.972
Subscale: Effort		
Mean	3.81	9.796
S.D.	2.768	5.940
Shapiro-Wilk Normality Test		
W	0.751	0.875
p-value	0.02	0.227
Subscale: Frustration		
Mean	3.537	7.755
S.D.	2.54	4.962
Shapiro-Wilk Normality Test		
W	0.845	0.823
p-value	0.125	0.080

Performance $t(6) = 1.456$, $p = .196$. The repeated measures *t*-test for the *Frustration* subscale showed there was a significant effect of the interaction modality on *Frustration*, $t(6) = 2.53$, $p = .04$. This implied that interacting with the robot remotely was more frustrating than in-person interaction. The results of the Wilcoxon Signed-Rank test showed that the interaction modality had an effect on *Effort* ($Z = -2.154, p = 0.031$). This implies that interacting with the robot remotely required more effort than interacting with the robot in-person.
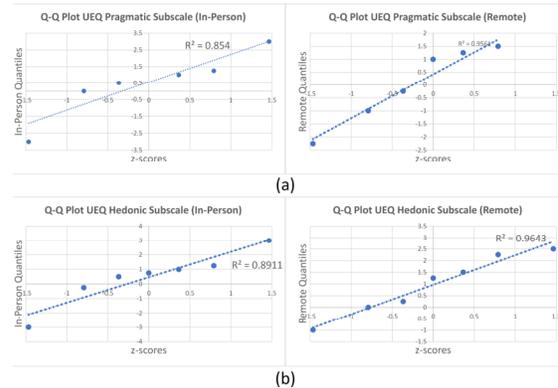
Figure 4: Q-Q Plot of UEQ (a) *Pragmatic* subscale and; (b) *Hedonic* subscale showing that the data seems to follow a normal distribution.

4.5. UEQ

The Q-Q Plots appear to indicate that the data for the UEQ subscales were normally distributed, Figure 4. Both subscales passed the Shapiro-Wilk normality test,

see Table 4, which meant we could use the *t*-test. For the *Pragmatic* subscale, the repeated measures *t*-test *t*(6) = -0.254, *p* = .81 showed there was no effect of the interaction modality on the pragmatic quality of the experience. For the *Hedonic* subscale the repeated measure *t*-test *t*(6) = .950, *p* = .38, showed that there was no effect by the interaction mode on the hedonic quality of the experience.

Table 4: Statistics and Normality Test Results UEQ

	In-Person (*N*=7)	Remotely (*N*=7)
Subscale: Pragmatic		
Mean	0.464	0.25
S.D.	1.805	1.458
Shapiro-Wilk Normality Test		
W	0.897	0.857
p-value	0.351	0.157
Subscale: Hedonic		
Mean	0.464	0.96
S.D.	1.82	1.27
Shapiro-Wilk Normality Test		
W	0.912	0.954
p-value	0.468	0.869

5. Discussion

In this paper we presented a pilot study of a Remote-HRI methodology in which we examined whether:

1. The participants' mode of interaction (in-person versus remotely) with the robot affected their attitudes towards the robot;
2. The workload associated with interacting with the robot depended on the interaction mode and;
3. The quality of the participants' experience with the robot was affected by the interaction mode.

To address these questions, we first examined participants' attitudes towards robots using the NARS and RoSAS inventories and found that the mode of interaction had no effect. We then used the TAM2 inventory to examine the acceptance of social robot technology and the intention to adopt such technology. We also found that the interaction mode had no effect on these attitudes.

The interaction workload was examined using the NASA TLX scale and we found that the interaction mode affected the level of effort and frustration reported, specifically, that interacting with the robot remotely required more effort and triggered more frustration than the in-person interaction. This is an important finding, and we believe that this merits further investigation on two fronts: (i) the voice recognition capabilities of the robot and; (ii) the technological infrastructure of the experiment environment. Regarding the first issue, the Zenbo robot showed difficulty in recognizing the wake-up command "Hey Zenbo" in the in-person interaction mode with participants sometimes shouting or changing their pronunciation in an effort to get the robot to respond. This effect was noticed for most of the other nine commands. We suspect that this may be a result of the dataset used to train the robot's voice recognition system combined with the fact that the experiment was performed at a Caribbean university where the participants spoke with accents from three different Caribbean countries. The second issue is simpler to rectify, and that is to ensure a reliable wireless or wired network connection, and a speaker that produces a clear sound so that the robot can hear and process the remote voice commands. We observed that when there was slight distortion in the speaker because a participant was speaking too loudly into the remote device, the Zenbo robot could not process the command.

We investigated the quality of the participants' experience using the Short UEQ inventory and found that the experiences were neutral and that there was no significant difference in the quality of experience based on the interaction mode. We suspect that the negative experience may also be related to the Zenbo voice recognition system since the majority of the commands were voice commands and Zenbo had demonstrable difficulty in responding to them.

5.1. Limitations

These results have limitations beyond the inconsistent response of the robot due to its voice recognition system. First, the extent to which command repetition affected effort and frustration cannot be quantified since we did not gather that data and only observed this phenomenon as part of the experiment. Second, the effect of the supporting technology was not examined since we focused on other aspects of the Remote-HRI methodology. The quality of the network connection may have influenced aspects of the interaction since there were some points when performance degraded and affected sound and/or video quality. These factors need to be captured so that the remote environment can be adequately designed and its effects mitigated. Lastly, we did not capture the effect of the smartphone remote device versus the laptop remote device and whether that affected any of our measures. We also must examine this in our full study.

6. Conclusions and Future Work

The ability to conduct HRI studies has been constrained by the current COVID-19 pandemic.

Finding a way to conduct these studies in an ethical and safe way is important, especially since most HRI research focuses on using social robots to improve quality of life for two of the most at-risk populations – the elderly and those with underlying health conditions.

In this paper we presented the results of a pilot study designed to assess the possibility of a Remote-HRI methodology. The results indicated that for our first and third questions, the participants' perceptions of the robot and their user experiences were unaffected by interaction modes. This may imply that experiments focused on evaluating user attitudes towards robots can be performed remotely once the commands can be issued vocally with limited physical contact. The results for our second question require further analysis due to two other factors that were not examined as part of this study: the supporting technology and the robot's voice recognition system. We do not believe that these two factors would prevent adoption of Remote-HRI, indeed, we are confident that upon fuller study, we will be able to provide mitigation strategies to reduce the effects of these factors.

Consequently, we will undertake future work in a full study to validate these results and also to investigate whether the participants' age and dialect has some effect on attitudes and experiences within the interaction modes. We will also evaluate whether different types of robotic applications are more suited for different types of Remote-HRI interaction besides those that are primarily voice driven, since successfully facilitating physical interactions will have a significant impact on the feasibility of R-HRI studies.

7. References

[1] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams, 'Where to Next? The Impact of COVID-19 on Human-Robot Interaction Research', *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 1, pp. 1–7, Jul. 2020, doi: 10.1145/3405450.

[2] H. Robinson, B. MacDonald, N. Kerse, and E. Broadbent, 'The psychosocial effects of a companion robot: a randomized controlled trial', *Journal of the American Medical Directors Association*, vol. 14, no. 9, pp. 661–667, 2013.

[3] E. Martinez-Martin and A. P. del Pobil, 'Personal robot assistants for elderly care: an overview', in *Personal Assistants: Emerging Computational Technologies*, Springer, 2018, pp. 77–91.

[4] S. Oh and Y. H. Oh, 'Understanding the Preference of the Elderly for Companion Robot Design', in *International Conference on Applied Human Factors and Ergonomics*, 2019, pp. 92–103.

[5] M. A. Al-Taee, R. Kapoor, C. Garrett, and P. Choudhary, 'Acceptability of robot assistant in management of type 1 diabetes in children', *Diabetes Technology & Therapeutics*, vol. 18, no. 9, pp. 551–554, 2016.

[6] O. A. Blanson Henkemans, S. Van der Pal, I. Werner, M. A. Neerincx, and R. Looije, 'Learning with Charlie: a robot buddy for children with diabetes', in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 406–406.

[7] S. Schneider and F. Kummert, 'Comparing the effects of social robots and virtual agents on exercising motivation', in *International Conference on Social Robotics*, 2018, pp. 451–461.

[8] A. N. Barata, 'Social Robots as a Complementary Therapy in Chronic, Progressive Diseases', in *Robotics in Healthcare*, Springer, 2019, pp. 95–102.

[9] CDC, 'Coronavirus Disease 2019 (COVID-19)', *Centers for Disease Control and Prevention*, Feb. 11, 2020. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html (accessed Jul. 24, 2020).

[10] M. Hammontree, P. Weiler, and N. Nayak, 'Remote usability testing', *interactions*, vol. 1, no. 3, pp. 21–25, 1994.

[11] K. E. Thompson, E. P. Rozanski, and A. R. Haake, 'Here, there, anywhere: remote usability testing that works', in *Proceedings of the 5th conference on Information technology education*, 2004, pp. 132–137.

[12] J. Rubin and D. Chisnell, *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons, 2008.

[13] H. R. Hartson, J. C. Castillo, J. Kelso, and W. C. Neale, 'Remote evaluation: the network as an extension of the usability laboratory', in *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*, Vancouver, British Columbia, Canada, 1996, pp. 228–235, doi: 10.1145/238386.238511.

[14] M. Hertzum, P. Borlund, and K. B. Kristoffersen, 'What Do Thinking-Aloud Participants Say? A Comparison of Moderated and Unmoderated Usability Sessions', *International Journal of Human–Computer Interaction*, vol. 31, no. 9, pp. 557–570, Sep. 2015, doi: 10.1080/10447318.2015.1065691.

[15] A. Vasalou, B. D. Ng, P. Wiemer-Hastings, and L. Oshlyansky, 'Human-moderated remote user testing: Protocols and applications', in *8th ERCIM Workshop, User Interfaces for All, Wien, Austria*, 2004, vol. 19.

[16] M. S. Andreasen, H. V. Nielsen, S. O. Schrøder, and J. Stage, 'What happened to remote usability testing? an empirical study of three methods', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, Apr. 2007, pp. 1405–1414, doi: 10.1145/1240624.1240838.

[17] K. Chalil Madathil and J. S. Greenstein, 'An investigation of the efficacy of collaborative virtual reality systems for moderated remote usability testing', *Applied Ergonomics*, vol. 65, pp. 501–514, Nov. 2017, doi: 10.1016/j.apergo.2017.02.011.

[18] World Health Organization and others, 'Transmission of SARS-CoV-2: implications for infection prevention

precautions: scientific brief, 09 July 2020', World Health Organization, 2020.

[19] A. Huber and A. Weiss, 'Developing Human-Robot Interaction for an Industry 4.0 Robot: How Industry Workers Helped to Improve Remote-HRI to Physical-HRI', in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna, Austria, Mar. 2017, pp. 137–138, doi: 10.1145/3029798.3038346.

[20] B. A. Mackey, P. A. Bremner, and M. Giuliani, 'Immersive Control of a Robot Surrogate for Users in Palliative Care', in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge, United Kingdom, Mar. 2020, pp. 585–587, doi: 10.1145/3371382.3377445.

[21] S. D. Sierra, M. F. Jiménez, M. C. Múnera, A. Frizera-Neto, and C. A. Cifuentes, 'Remote-operated multimodal interface for therapists during walker-assisted gait rehabilitation: a preliminary assessment', in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, Daegu, Republic of Korea, Mar. 2019, pp. 528–529, Accessed: Jul. 16, 2020. [Online].

[22] K. Qian, J. Niu, and H. Yang, 'Developing a gesture based remote human-robot interaction system using kinect', *International Journal of Smart Home*, vol. 7, no. 4, pp. 203–208, 2013.

[23] G. M. Nagy, J. E. Young, and J. E. Anderson, 'Are Tangibles Really Better? Keyboard and Joystick Outperform TUIs for Remote Robotic Locomotion Control', in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, Portland, Oregon, USA, Mar. 2015, pp. 41–42, doi: 10.1145/2701973.2701978.

[24] K. Stubbs, D. Wettergreen, and I. Nourbakhsh, 'Using a robot proxy to create common ground in exploration tasks', in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, Amsterdam, The Netherlands, Mar. 2008, pp. 375–382, doi: 10.1145/1349822.1349871.

[25] J. Burke and R. Murphy, 'RSVP: an investigation of remote shared visual presence as common ground for human-robot teams', in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, Arlington, Virginia, USA, Mar. 2007, pp. 161–168, doi: 10.1145/1228716.1228738.

[26] Z. Prasov, 'Shared gaze in remote spoken hri during distributed military operation', in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, Boston, Massachusetts, USA, Mar. 2012, pp. 211–212, doi: 10.1145/2157689.2157760.

[27] S. Honig and T. Oron-Gilad, 'Comparing Laboratory User Studies and Video-Enhanced Web Surveys for Eliciting User Gestures in Human-Robot Interactions', in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge, United Kingdom, Mar. 2020, pp. 248–250, doi: 10.1145/3371382.3378325.

[28] J. Mumm and B. Mutlu, 'Human-robot proxemics: physical and psychological distancing in human-robot interaction', in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 331–338.

[29] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, D. S. Syrdal, and C. L. Nehaniv, 'An empirical framework for human-robot proxemics', *Procs of new frontiers in human-robot interaction*, 2009.

[30] R. Macefield, 'How to specify the participant group size for usability studies: a practitioner's guide', *Journal of Usability Studies*, vol. 5, no. 1, pp. 34–45, 2009.

[31] S. Byrne, 'Zenbo family robot destroys Computex with cuteness before it even begins', *CNET*. https://www.cnet.com/news/zenbo-family-robot-destroys-computex-with-cuteness-before-it-even-begins/ (accessed Jul. 26, 2020).

[32] 'Zenbo | Intelligent Robot', *ASUS Global*. https://www.asus.com/Commercial-Intelligent-Robot/Zenbo/ (accessed Jul. 26, 2020).

[33] S.-E. Chien *et al.*, 'Age Difference in Perceived Ease of Use, Curiosity, and Implicit Negative Attitude toward Robots', *J. Hum.-Robot Interact.*, vol. 8, no. 2, p. 9:1–9:19, Jun. 2019, doi: 10.1145/3311788.

[34] 'Jitsi.org - develop and deploy full-featured video conferencing', *Jitsi*. https://jitsi.org/ (accessed Jul. 16, 2020).

[35] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, 'Altered attitudes of people toward robots: Investigation through the Negative Attitudes toward Robots Scale', in *Proc. AAAI-06 workshop on human implications of human-robot interaction*, 2006, vol. 2006, pp. 29–35.

[36] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, 'The robotic social attributes scale (RoSAS) development and validation', in *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, 2017, pp. 254–262.

[37] V. Venkatesh and F. D. Davis, 'A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies', *Management Science*, vol. 46, no. 2, pp. 186–204, Feb. 2000, doi: 10.1287/mnsc.46.2.186.11926.

[38] S. G. Hart, 'NASA-task load index (NASA-TLX); 20 years later', in *Proceedings of the human factors and ergonomics society annual meeting*, 2006, vol. 50, no. 9, pp. 904–908.

[39] B. Laugwitz, T. Held, and M. Schrepp, 'Construction and Evaluation of a User Experience Questionnaire', in *HCI and Usability for Education and Work*, Berlin, Heidelberg, 2008, pp. 63–76, doi: 10.1007/978-3-540-89350-9_6.

[40] A. Ghasemi and S. Zahediasl, 'Normality Tests for Statistical Analysis: A Guide for Non-Statisticians', *Int J Endocrinol Metab*, vol. 10, no. 2, pp. 486–489, 2012, doi: 10.5812/ijem.3505.

[41] J. C. De Winter, 'Using the Student's t-test with extremely small sample sizes', *Practical Assessment, Research, and Evaluation*, vol. 18, no. 1, p. 10, 2013, doi: https://doi.org/10.7275/e4r6-dj05.