DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2019

Investigating Augmented Reality for Improving Child-Robot Interaction

EMMELI HANSSON

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Abstract

Communication in HRI, both verbal and non-verbal, can be hard for a robot to interpret and to convey which can lead to misinterpretations by both the human and the robot. In this thesis we look at answering the question if AR can be used to improve communication of a social robot's intentions when interacting with children. We looked at behaviors such as getting children to pick up a cube, place a cube, give the cube to another child, tap the cube and shake the cube. We found that picking the cube was the most successful and reliable behavior and that most behaviors were slightly better with AR.

Additionally, endorsement behavior was found to be necessary to engage the children, however, it needs to be quicker, more responsive and clearer. In conclusion, there is potential for using AR to improve the intent communication of a robot, but in many cases, the robot behavior alone was already quite clear. A larger study would need to be conducted to further explore this.

Sammanfattning

I Människa-Robot Interaktion kan både verbal och icke-verbal kommunikation vara svårt för en robot att förstå och förmedla vilket kan leda till missförstånd från både människans och robotens håll. I den här rapporten vill vi svara på frågan ifall AR kan användas för att förbättra kommunikationen av en social robots avsikter när den interagerar med barn. De beteenden vi kollade på var att få ett barn att plocka upp en kub, placera den, ge den till ett annat barn, knacka på den och skaka den. Resultaten var att plocka upp kuben var det mest framgångsrika och pålitliga beteendet och att de flesta beteenden var marginellt bättre med AR.

Utöver det hittade vi också att bifalls beteenden behövdes för att engagera barnen men behövde vara snabbare, mer responsiva och tydligare. Sammanfattningsvis finns det potential för att använda AR, men i många fall var enbart robotens beteenden redan väldigt tydliga. En större studie skulle behövas för att utforska detta ytterligare.

Investigating Augmented Reality for Improving Child-Robot Interaction

Emmeli Hansson
emmelih@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

Communication in HRI, both verbal and non-verbal, can be hard for a robot to interpret and to convey which can lead to misinterpretations by both the human and the robot. In this thesis we look at answering the question if AR can be used to improve communication of a social robot's intentions when interacting with children. We looked at behaviors such as getting children to pick up a cube, place a cube, give the cube to another child, tap the cube and shake the cube. We found that picking the cube was the most successful and reliable behavior and that most behaviors were slightly better with AR.

Additionally, endorsement behavior was found to be necessary to engage the children, however, it needs to be quicker, more responsive and clearer. In conclusion, there is potential for using AR to improve the intent communication of a robot, but in many cases, the robot behavior alone was already quite clear. A larger study would need to be conducted to further explore this.

KEYWORDS

HRI, CRI, AR, Social Robotics, Robot behavior

1 INTRODUCTION

Communication in Human-Robot Interaction (HRI) is not always smooth. One reason for this is that communication between humans make use of many minute cues such as prosody in speech, facial expressions and body language which can be hard or impossible for a robot to interpret. This can lead to misinterpretation, both by the human and the robot. While speech recognition and facial tracking in HRI are improving there is still a long way to go for it to be completely robust.

So why is this a problem? In social robotics, it is important that the robot can navigate social situations, and if those social situations rely on non-verbal cues for example, the robot needs to be able to understand those too. On the other hand, it is just as important for the robot's intentions to be clear to the human counterpart for a smooth interaction. One way of tackling that is to use speech, or imitating other human behaviors, especially since humans also use nonverbal behavior to communicate their intentions. However, does this mean that a robot can not convey its intentions in a reliable way without speech or with limited non-verbal capacities like facial expressions or body language if they don't have human-like embodiments? Are there other reliable ways of communicating, less dependent on closely imitating humans? Social robots such as the company Anki's Cozmo, Figure 1, which is not humanoid, is still able to convey emotion and has a wide range of possible behaviors that can be played out by the robot. Cozmo has a screen with animated eyes, which conveys a version of facial expressions based in animatronics. Another avenue to explore for discovering novel ways of conveying intention could be the use of Augmented Reality (AR). If one were to combine AR and robot behaviors like the ones used by Cozmo, maybe the behaviors could be enhanced and the robot's intentions made clearer.

Making the communication between human and social robot less error-prone, less reliant on speech or facial recognition, and less limited by a social robot's possible lack of human-like embodiments, would increase the usefulness of social robots and increase the kind of situations where they could be used. Examples are situations where human resources are not enough and it would be beneficial to bridge that gap with robots, or in situations where there might be need of monitoring but a human monitor risks interfering. Or in situations were the persons interacting with the robot might not speak the same language as each other or as the robot and social HRI could be used to minimize language barriers. Meaning that, if the robot interaction does not rely on speech to be understood, the participants do not need to speak the same language but will hopefully all still understand what the robot wants.

An interesting area to study within HRI is child-robot interaction (CRI), to get insight into how children might perceive and receive a social robot. The use of social HRI has been studied in the context of interacting with children in hospital settings [3] or simply in childcare [16]. These are also situations were the interaction and communication would benefit from being reliable.

This paper contributes to the research area by studying the potential of AR to improve non-verbal communication in child-robot interaction.

Figure 1: Anki's Cozmo, a palm-sized social robot

2 RELATED WORK

Exploring how to communicate a robot's intentions clearer is an important step in making HRI more reliable, and there are many studies that already study the potential of AR in that area. First, however, it is necessary to define what is meant by AR in this paper.

Master's Thesis, May 2019, Stockholm, Sweden
2019.

2.1 Different types of AR

Augmented Reality (AR) is a kind of Mixed Reality (MR), which is the combination of the real world and the virtual world, in contrast to Virtual Reality (VR) in which the user is entirely immersed in the virtual world [9]. AR can be defined as specifically the cases where virtual objects, for example computer graphics, are added to a real-world environment [9]. This can be accomplished today through the use of Head-Mounted Displays (ARHDM), like the Microsoft Hololens, or hand-held displays like smartphones. A third version of AR is where the "augmented" part is projected onto a real-world space through the use of a projector, as in [1, 5, 6, 10, 13, 14]. This is sometimes called Spatial Augmented Reality (SAR) [6]. Combining SAR and HRI opens new doors for AR in HRI, to explore the different kinds of interactions that present themselves [6]. However, the definition of these terms seems to be fluid. For example Sibirtseva et al. [14] use the term Mixed Reality for the use of a ARHMD in their user-tests, and AR for the version of their user-tests using a projector to project augmented reality elements onto a surface.

2.2 Communicating robot intention

In human-human communication, one might use pointing or arm-motions to communicate intent. In HCI situations, the development of MR and AR is opening new ways for systems to visualize the robot intentions to the user [11, 18]. Clear robot intent communication can be seen as a requirement for human-robot collaboration to work smoothly [7, 11] . One way to ensure this is to make sure the robot's motions are predictable, or preferably legible to the user[7], but the use of AR would mean that there are alternatives for understanding the robot's intention that do not rely on robot motion. Although, the AR still needs to be legible and easily understood by the user.

However, what is meant by wanting to communicate the robot's intent? What could the robot want a human to know? Walker et al. [17] study how a robot can potentially communicate the direction it is going, using a flying robot and a ARHMD on the user. They tested an assembly task in a warehouse, where the user and the robot would share space and resources. The user would see one of four designs in the ARHMD. One of their designs, consisting of virtual imagery overlayed on the environment, showed the robot's planned flight as lines and navigation waypoints as spheres in the 3D space and was found to have the best results for efficiency and for the perception of the robot as a personal work partner. However, they question the scalability of that design. Walker et al. also talk about the limitations of projector systems as an alternative to ARHMD, such as occlusion of the projected image by either the robot or the user. [17]

Sibirtseva et al. [14] tested an object retrieval task, where the input from the user was verbal requests and the output from the robot was one of three visualization techniques highlighting what object to retrieve. One condition was with MR, which in their case was AR with a head mounted display, and another condition was displaying the scene with highlights on a side monitor. The third condition was with AR using a projector. The intent of the robot was to show the user what object it thought the user wanted it to pick and get confirmation on if it was the correct one. All conditions had the same error rate, however, the AR condition was deemed the least disruptive and most likely to be used again. [14]

According to Sibirtseva et al. [14], AR using a projector is a good strategy. Another case using AR and a projector is where Andersen et al. [1] look at how using a projection of intention for human-robot collaborative tasks in industrial settings can lessen the difficulty for humans to predict the robot intent. In these industrial settings, it is increasingly useful with more intelligent robots that can share workspace, either by just having tasks in the same space as the human workers or collaborating on a task. Their results are that projection increases the usability of such robots and humans perception of them, especially for user satisfaction and usability. Results are not conclusive, but suggest the potential of projecting AR. [1]

Another shared space case where AR using projection is used is by Chadalavada et al. [5] where the authors look at a mobile autonomous fork-lift robot and how the projection of its intentions or path trajectory onto the floor directly in front of it is perceived by a human sharing the same space or moving towards it. Their pilot studies show that there is a significant positive increase in the human's perception of the robot when it projects its intentions on the floor, compared to when it does not project its intentions. Specifically, the robot's intentions and the robot itself were perceived as more predictable and transparent [5], which means that the intentions of the robot were clearer to the human worker.

In conclusion, the use of some kind of AR imagery can make the robot's intentions clearer. The examples from the previous research pertain to industrial settings where the robot's intentions being clear is critical for a safe environment. Nonetheless, they indicate ways of showing robot intention and these ways could potentially be applied in different situations with different kinds of robot intentions. Additionally, all these studies were done with adults, and we will look at how the use of AR can be applied to robot interactions with children. Because this study will be done on a group of children, the type of AR used will be SAR, using a projector directed at a table, but we will refer to it simply as AR. The reason for this choice is that using ARHDM for more than one child at a time was not feasible for this study and since SAR has successfully been used for conveying intention that version of AR was chosen instead.

2.3 The use of color in robotics

One aspect that AR adds to the communication of intent is the possibility to use colors as part of the augmentation.

Song and Yamada [15], using colored lights on a robot, found that a relaxed state could be conveyed with white, sadness with blue. Happiness was hard to convey and angry was conveyed with red but deemed better with more modalities. Overall, emotions were conveyed better with more than one modality [15].

According to a second study by Löffler et al. [8], joy is best conveyed via color and motion. They used Hue, Saturation, Brightness (HSB) values for the color coding, specifically some kind of warm yellow for joy (HSB:45/100/100). Sadness was better conveyed via sound, fear via motion while anger also worked the best via color, specifically red (HSB:0/100/100). They used a robot displaying colored light and sound, designed without anthropomorphic, zoomorphic or too many mechanical elements. The colors used for sadness and fear, and rejected by Löffler et al. [8], were dark blue-gray (HSB:230/40/40) and dark gray (HSB:0/0/20) respectively.

In conclusion, there have been mixed findings in previous research on the use of colors in HRI, but these studies still give an indication of how it could be used. Both studies show that red can be used to convey anger [8, 15], however, for ethical reasons studies with children should be careful of using strong negative feedback, and we decided not to use any strong negative feedback at all. The more interesting colors were, therefore, the ones used for sadness, joy and a relaxed state. Blue for sadness and yellow for joy did not convey well with projected AR, however white for relaxed worked well and was deemed applicable for the study at hand.

2.4 Child-robot interaction

Overall, previous research show that children treat a social robot as a social agent and are inclined to interact with it [3, 4, 16].

However, in child-robot interaction (CRI), children are a special user group and can not be treated as adults. Children have a lot of imagination and that is reflected in their interactions with robots. This also enables the exploration of different and engaging social interactions for CRI. [3].

Beran et al. [4] looks at how children perceive robots and if they attribute human traits to them. Their results show that they do, one explanation being that the children connected the robot to what they knew about the world around them, specifically related to human actions or actions perceived to be human. Additionally, based on their results they propose that it may not be necessary to make the robot perfectly consistent because children understand that to some extent humans or pets are not. Also, children might be naturally inclined to see the robot in a good light and with curiosity and therefore want to interact and include the robot in their play. [4]

Belpaeme et al. [3] found that children react best to robots that adapt their behavior, to a robot with flexible behavior or else the child can quickly lose interest.

Shahid et al. [12] specifically looks at CRI for children across cultures and how their treatment and reaction to a social robot differs. In their study, they found that Pakistani and younger children liked interacting with their social robot more than the Dutch children and older children. So they found that age and culture were important factors, and underlying that social norms should be part of the robot design process when applicable.

In conclusion, in CRI, it is important to develop the robot interaction with that user group in mind and adapt the robot behavior and intent communication thereafter.

2.5 Problem statement

Based on the usefulness of AR for communicating intention, and the potential and complexity of CRI, the following research question was developed:

- Can AR improve the communication of a social robot's intentions when interacting with children?

3 METHOD

To answer the research question stated above, a pilot experiment was developed. The objective of the pilot experiment was to determine which actions done by Cozmo were expressing its intentions in the most accurate manner. In order to do so, the robot was preprogrammed with several behaviors. The behaviors were selected to convey a specific intention to the children. Then, depending on their reaction, the different behaviors were evaluated on their effectiveness to infer the robot's desired intention.

The experiment was run in two different conditions: "Robot only" (OR) and "With AR" (AR). In both conditions, the robot would try to show its intentions mainly by means of movement of its body and lifting its forklift-like arms. Cozmo's facial expression mainly consisted of two big animated eyes on a small display. The animations were preprogrammed to the behaviors by Anki, so they were not customizable and were the same between conditions for the same behaviors.

In our experimental condition, "With AR", Augmented Reality was added to the expressive modalities of Cozmo. AR elements supported the intentions that Cozmo wanted to convey, meaning that in this condition the behaviors were done by combining Cozmo's movements with AR elements.

3.1 Cozmo

Cozmo is a palm-sized social robot, see Figure 1, with the dimensions 6.4 cm wide, 7.6 cm long and 10.1 cm high. For size reference, look at Figure 2 of two children interacting with Cozmo. The robot has treads for moving forward and backwards, and a square face with a screen on which its eyes are animated. Cozmo is a very expressive and sociable robot by design, through the eyes on the screen, the sounds it makes and the way it is preprogrammed to move. It does not have arms but has a kind of forklift that it uses as arms. On the forklift-like arms there are small spikes that can be used to lift cubes that come with the product.

Cozmo has preprogrammed behaviors and games that it can do that are controlled through an application on a tablet. However, it is also customizable and can be programmed through the use of a Python SDK, which is what we used. We chose not to use the

Figure 2: A group of two children interacting with Cozmo

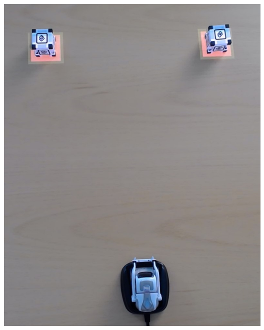

Figure 3: Set-up for Pick

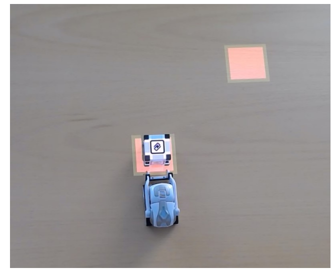

Figure 4: Set-up for Pick and Place

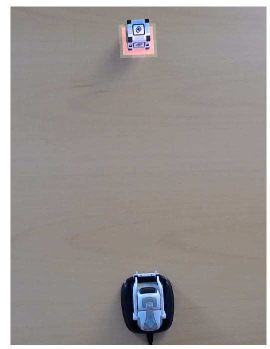

Figure 5: Set-up for "Pass Responsibility", Tap, Shake

sounds, to test fewer modalities at a time. Additionally, Cozmo has more complicated functionalities, such as facial recognition, that we chose not to use in order to simplify the experiment and to make it easier for a wizard to control Cozmo during the experiments. For more information and example videos of Cozmo in action, one can look at Anki's website [2].

3.2 Experiment Protocol

The study was conducted in an elementary school in Stockholm during their after-school program ("fritids" in Swedish). The children were in the age range of 8-10 (M=9, SD=0,4) years old. 39 children participated, divided into 15 groups by their teacher, 9 groups of 3 and 6 groups of 2. Ideally would have been to have only groups of three but that was not possible logistically. There were 7 groups in the OR condition and 8 groups in the AR condition. The experiment was divided into different episodes, where a specific intention was assessed in a given set-up in each episode. All the behaviors were followed by a posterior feedback response from the robot that would endorse the children's actions regardless of whether or not they understood what the robot was conveying to them.

3.2.1 Intention in this context. An intent is an aim or a goal. An instruction is a directive. In the experiment design Cozmo wants the children to do certain actions, which could be considered as Cozmo trying to give them instructions. However, the purpose of the experiment is for the children and Cozmo to achieve a goal together, through the children understanding what the robot is trying to convey and Cozmo being responsive to their actions and acting accordingly. Hence, like in the studies using AR to convey robot intention [1, 5, 14, 17], Cozmo is still mainly trying to show the aim of its actions. Reaching that goal is a joint effort, therefore we will be using the terms intent and intention.

3.2.2 Set-ups. The general set-up consisted of the robot with one or two cubes together with the projector and camera pointing down to the area where the robot was. The cubes were used by the kids to perform the actions that Cozmo was communicating to them. Two different set-ups were used and they did not vary between the conditions.

(1) **Set-up with 1 cube:** There was only one cube on the table, which initially was aligned with the robot in the center. The distance between the cube and the robot could vary depending on the given episode that was being run. See Figures 3 and 4.
(2) **Set-up with 2 cubes:** There were two cubes on the table at the same distance to the robot, and each of them was aligned to the position of each child. See Figure 5.

3.2.3 Episodes. The experiment consisted of the six episodes listed below. In each episode, one specific intention was assessed, to see if the robot conveyed that one intention to the children (e.g. pick up one cube). The goal was for the children to explore, discuss and try to guess the intention and act accordingly. Each episode was run in the same way for all children independent of the condition, the only difference between the conditions being that in the AR condition AR elements were added.

(1) **Pick cube:** (Figure 3) The robot's intention was for one of the children to pick up one cube.
(2) **Pick cube:** (Figure 3) The robot's intention was for one of the children to pick up the other cube.
(3) **Pick and Place cube:** (Figure 4) The robot's intention was for one of the children to pick up the cube, and then place it in a specific location indicated by the robot.
(4) **Re-assign responsibility on pick cube ("Pass Responsibility"):** (Figure 5) The robot's intention was to convey the message "let your friend play as well". The robot starts with the pick activation from #1. When one of the children has picked up the cube, the robot's intention was to transfer the responsibility of picking the cube to the other child. The goal was for the cube to be placed back and for the other child to

Table 1: Details of Cozmo behavior and AR per episode

#	Episode	Cozmo behavior (OR and AR condition)	AR elements (AR condition)
1 & 2	Pick cube	Approach cube and move head up and down like staring at the cube and then the child. Then, take the cube and lift it up.	A white square over the cube to be picked.
3	Pick and Place cube	First do the pick behavior, then go to location to place it and tap on table	A white square over the cube to be picked, disappears when Cozmo is at new place where a white square appears in the spot the cube is to be placed. If child does not understand, an arrow appear pointing at the cube appears.
4	Pass responsibility	When first child picks the cube, Cozmo acts a little confused, turns to second child and indicates through motion and head movement that the second child should take the cube	A white square over the cube to be picked which disappears when child takes the cube/when Cozmo gets confused. To re-assign responsibility an arrow pointing at the another child appears.
5	Tap cube	Goes to the cube and makes a tapping motion, if the children do not understand go to the cube and tap it	A smaller white square/spotlight on the cube (only on the cube) appears, blinking "rapidly". Blinking should only start after Cozmos starts tapping motion
6	Shake cube	Does the pick behavior and when the child has picked the cube, moves a little from side to side like "shaking" to show what he wants them to do with the cube.	A white square on the cube that Cozmo wants the child to pick. When the child picks the cube, the spot changes into a 3D cube that "shakes".

pick the cube. For the groups with three children, the robot behavior was repeated to get the third child to receive the cube.

(5) **Tap cube:** (Figure 5) The robot's intention was for one of the children to tap on the cube.
(6) **Shake cube:** (Figure 5) The robot's intention was for one of the children to shake the cube.

More detailed descriptions of the Cozmo behavior and the AR for each episode can be found in Table 1. The color white was used for the AR projections because white can be seen as a relaxing color in HRI, according to Song and Yamada [15]. The experiment will make use of Anki's Cozmo robot, Figure 1 , and a projector for the AR. The choice of AR was based on, projecting AR being used for intent communication in several previous studies [1, 5, 14], but specifically Sibirtseva et al. [14] used a projection type of AR for picking objects which is one of the tasks that will be used in the study. Sibirtseva et al. [14] highlighted specific objects with AR in their study, which we emulated. The arrows used in Episode 3 and 4 were inspired by Walker et al. [17] where similar imagery was used to show directional intent. The AR elements for Tap and Shake were harder to find inspiration for in related research, so they were brainstormed specifically for this study. A few versions were found and tested on ourselves to then choose the ones used. In the end, the Tap robot motion was based on one of the preprogrammed games where Cozmo approached a cube and the tapping consisted of lifting its forklift-like arm and bringing it down onto the cube. Lastly, the AR part of the method was developed in Unity using ARToolKit.

3.2.4 Procedure. Two or three children sat on chairs at the table, see Figure 2, and Cozmo knew their rough position. Experimenter 1 welcomed the children, told them to sit and gave them the instructions of the experiment. Other experimenters, one or two wizards sat in the room out of the children's sight, controlling the Cozmo behavior and AR. During the experiment, experimenter 1 tried to show as little verbal or facial expressions as possible that could give feedback to the children, by for example looking at their phone or notebook. In each episode, if the intended child behavior was not activated, the test could be run again, but each try lasted a maximum of 2 minutes, preferably 1 minute, with a maximum of 2 tries. For retrying, experimenter 1 placed Cozmo back in the start position and the wizard(s) restarted the behavior. No feedback from the experimenter would be given to the children in between tries or episodes, apart from telling them they were going to do the next run. All experiments were videotaped and the results came from annotating the videos.

3.3 Video Coding

As mentioned every run of the experiment was videotaped, and the results were gathered from annotating the videos. For each experiment, each episode was looked at separately, so the following things were coded per episode:

- If the task was a success or failure, a success being coded as when the children were able to guess correctly the robot's intentions, and failure if they did not.
- If the success happened on the first or second try.
- How much time a success took.

OR group #	3	5	7	8	10	13	17
Episode 1	no	yes	no	yes	yes	yes	yes
Episode 2	no	yes	yes	yes	yes	yes	yes
Episode 3	no	yes	no	no	yes	yes	yes
Episode 4	no	yes	no	yes	yes	yes	yes
Episode 5	no	no	yes	no	no	no	yes
Episode 6	no	yes	no	yes	yes	yes	yes

Figure 6: Successes in OR condition

AR group #	2	4	6	9	11/12	14	15	16
Episode 1	yes	yes	yes	yes	yes	yes	yes	yes
Episode 2	yes	yes	yes	yes	yes	yes	yes	yes
Episode 3	yes	no	yes	yes	yes	yes	yes	yes
Episode 4	yes	no	yes	yes	yes	no	yes	yes
Episode 5	no	yes	no	no	no	no	no	no
Episode 6	yes	no	yes	yes	yes	yes	yes	no

Figure 7: Successes in AR condition

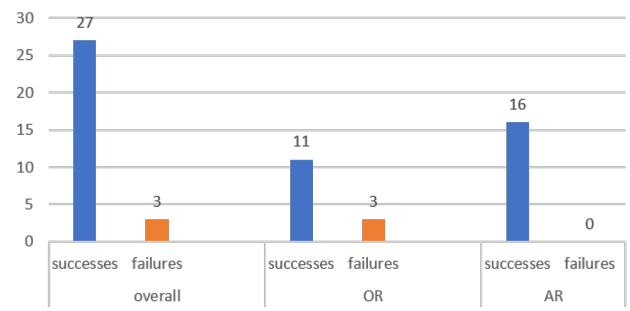

Figure 8: Successes and failures of Episode 1 and 2

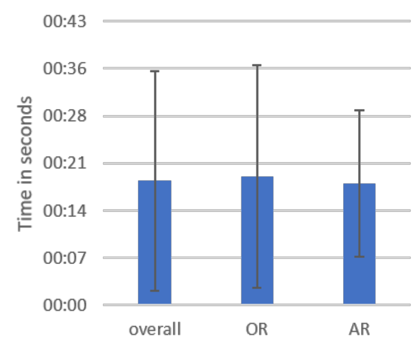

Figure 9: Average time of successes of Episode 1 and 2

We also noted overall comments about each group, specifically relating to the endorsement behavior, whether it seemed to be quick or slow and whether the children seemed to understand it.

There was a short interview with every group after the experiment, which was also video-taped, however, the results from these interviews were not very conclusive so they were not included in the results.

4 RESULTS

The first observation about the successes and failures of each episode over all the groups, from Figures 6 and 7, was that most episodes seem to have succeeded, especially for the AR groups. Only Episode 5 (Tap) had a high amount of failures. The two groups with the most failures were in the OR condition (groups 3 and 7), and one of them, group 3 failed all episodes. For group 3, none of the children dared handle the cubes. There was also only one group that succeeded on all episodes, group 17 in the OR condition. The most common result seems to be successes on all episodes except tapping, which was the case for 8 groups in total, where 3 were in the OR condition and 5 in the AR condition.

Overall 64% of the episodes in the OR condition succeeded to 36% that failed, and 75% succeeded in the AR condition versus 25% that failed.

4.1 Episode 1 and 2: Pick

Episodes 1 and 2 can be grouped together because they tested the same intent, which was to pick up a certain cube. The only difference between these two episodes was that the children were asked to pick up different cubes. By grouping the episodes together we get a total of 30 runs of the intent instead of 15. From Figure 8 one can see that most groups succeeded at picking up a cube, 27 successes against only 3 failures. Of those successes, 11 were in the OR condition, and 16 in the AR condition. Additionally, all the failures were in the OR condition, meaning that all AR runs of these episodes succeeded.

When looking at the time it took for each success, Figure 9, the overall average was 18 seconds (M=13sec, SD=16.41sec), with an average of 19 seconds (M=12sec, SD=16.57sec) for the OR condition, and an average of 18 seconds (M=12sec, SD=11.08sec) for the AR condition. Overall, the average times are very similar, the OR and AR condition only differing by 1 second, the AR condition being 1 second quicker. However, the standard deviation of the AR condition is significantly smaller. Consequently, the AR condition could be deemed more reliable, especially also since that condition had no failures. Although the success rate is very high over both conditions, so the AR condition is only slightly better.

4.2 Episode 3: Pick and Place

In Episode 3, where the children were to pick a cube and then place it in an area that Cozmo and/or the AR indicated, had overall 11 successes out of 15 groups. See Figure 10. Three of the failures were in the OR condition, making that 4 successes and 3 failures for the OR condition in this episode, while the AR condition had 7 successes and only 1 failure.

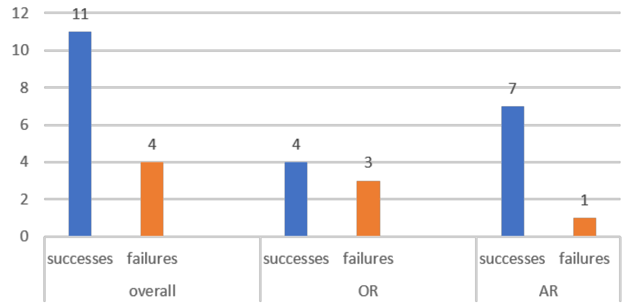

Figure 10: Successes and failures of Episode 3

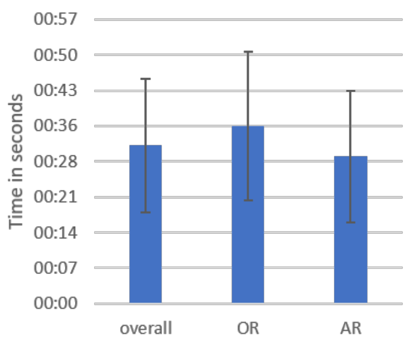

Figure 11: Average time of successes of Episode 3

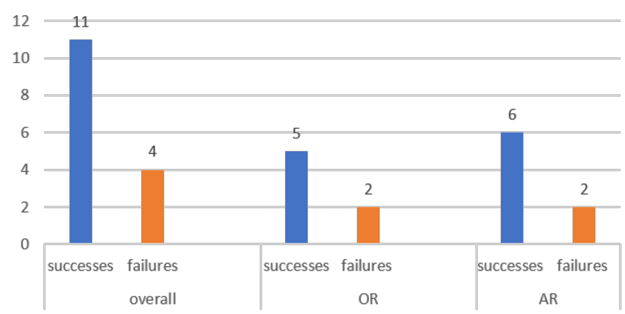

Figure 12: Successes and failures of Episode 4, first hand-over

Looking at the time of the successes in Episode 3 the overall average time was 32 seconds (M=30sec, SD=13.34sec), the average for the OR condition 36 seconds (M=42sec, SD=14.49sec) and the average for the AR condition 29 seconds (M=25sec, SD=13.21sec). The AR condition was slightly quicker than the OR condition. However, the times for the OR condition were based on only the 4 successes, while the AR condition times were based on 7 successes.

So, for picking and placing a cube AR seems to be the better choice.

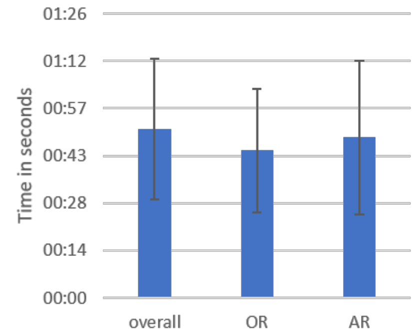

Figure 13: Average time of successes of Episode 4, first hand-over

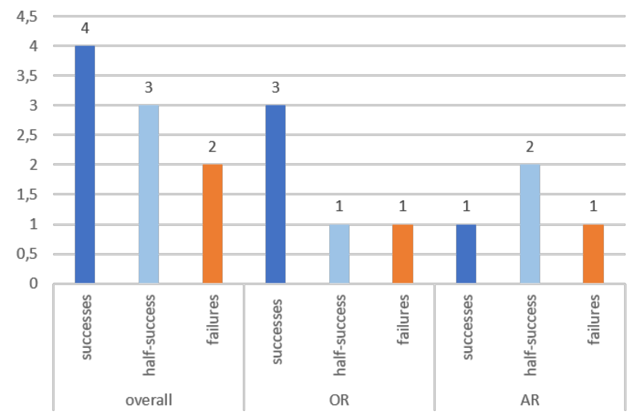

Figure 14: Successes and failures of Episode 4, second hand-over

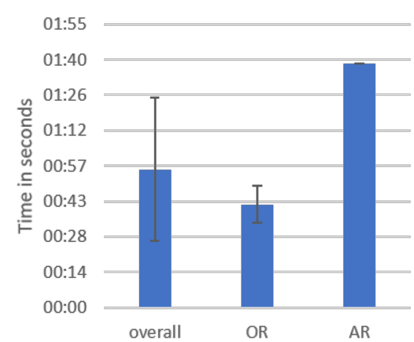

Figure 15: Average time of successes of Episode 4, second hand-over

4.3 Episode 4: Pass Responsibility

Episode 4 was when Cozmo was trying to get the children to share the cube, or to get possession of the cube to change child. For the results here we separated the groups of 2 children and the groups

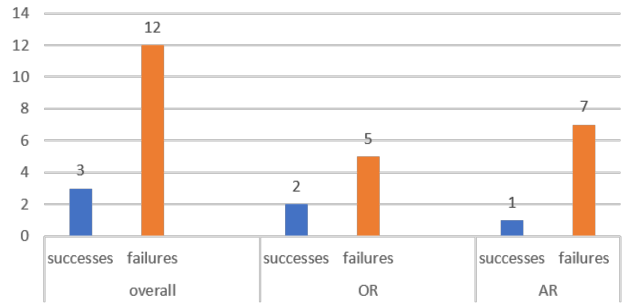

Figure 16: Successes and failures of Episode 5

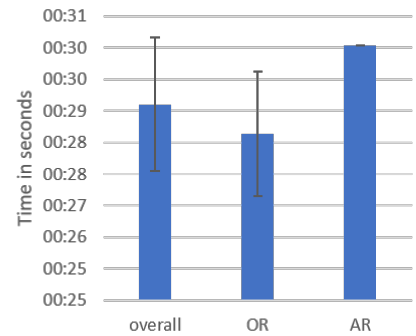

Figure 17: Average time of successes of Episode 5

of 3 children when looking at the second hand-over of the cube since that was only applicable to the groups of 3.

For the first hand-over, there were 11 successes overall and 4 failures, with 5 successes and 2 failures in the OR condition versus 6 successes and 2 failures in the AR condition, meaning the two conditions had very similar results. See Figure 12. The average time for these successes, see Figure 13 was overall 51 seconds (M=46sec, SD=21.24sec), 44 seconds (M=48sec, SD=18.43sec) for the OR condition and 48 seconds (M=45sec, SD=23.21sec) in the AR condition. It would seem that the OR condition has a slightly better time in this episode, for the first hand-over.

In the second hand-over, see Figure 14 which is applicable to the 9 groups of 3 children, there were 4 successes, 2 failures and 3 half-successes which is the cases when the first hand-over was a success but not the second. 3 of these successes could be attributed to the OR condition, with 1 half-success and 1 failure, while there was only one success in the AR condition, with 2 half-successes and 1 failure. It would seem that OR was better in this second hand-over as well. As for the times, seen in Figure 15 they were only measured on the 4 successes, but the average overall time was 56 seconds (M=46sec, SD=29.10sec), the average for the OR condition 42 seconds (M=43sec, SD=7.33sec) and the one success in the AR condition took 1 minute and 39 seconds. So according to those results OR would be a better choice for the "Pass Responsibility" scenario, however, the time measured for the second hand-over might not be reliable because of a too small number of groups measured.

4.4 Episode 5: Tap

This episode was by far the least successful, see Figure 16. There were 12 failures out of 15 groups and only 3 successes. 2 of these successes were in the OR condition while only 1 was in the AR condition. The time measurements could only be done on 3 groups, so they might not be accurate, but the overall average time was 29 seconds (M=30sec, SD=01:32sec), 29 seconds (M=29sec, SD=01:25sec) in the OR condition and 31 seconds for the one successful run in the AR condition. In conclusion, the tapping scenario did not work for the majority of the groups, neither with AR or without.

4.5 Episode 6: Shake

In Episode 6, the children were to pick up the cube and then shake it. Here we had 9 successes and 6 failures overall, see Figure 18. 5

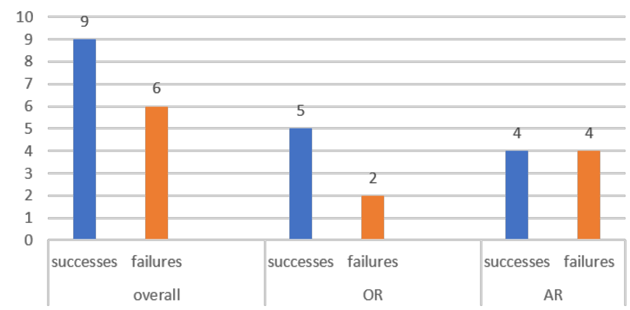

Figure 18: Successes and failures of Episode 6

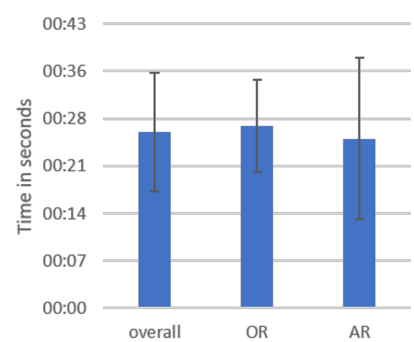

Figure 19: Average time of successes of Episode 6

successes and 2 failures in the OR condition, and 4 successes and 4 failures in the AR condition. Consequently, here the OR condition was slightly better than the AR condition.

Time-wise, it seems AR was the better choice. The average overall time was 26 seconds (M=28sec, SD=9.3sec), the average time for the OR condition was 27 seconds (M=29sec, SD=7.01sec) and the average time for the AR condition was 25 seconds (M=23sec, SD=12.15sec), see Figure 19.

This episode had a better success rate than Episode 5, however, it still had more failures than most of the other episodes.

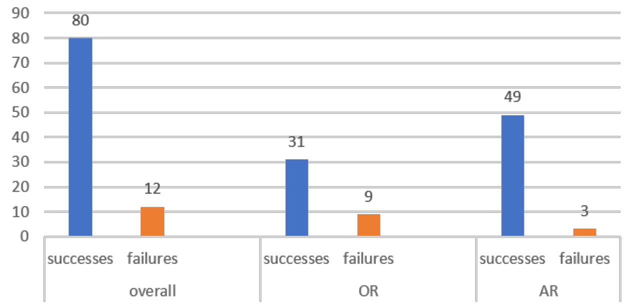

Figure 20: Successes and failures of pick over all the episodes

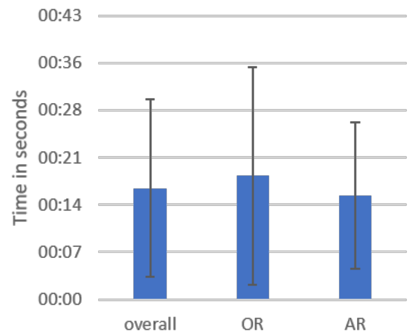

Figure 21: Average time of successes of pick over all the episodes

4.6 Pick over all the episodes

Another aspect to look at could be the pick behavior overall episodes since all episodes except Episode 5 were based on the children first picking up the cube. In some cases, when the episode was run more than once, the pick part might have succeeded in all the tries, while the actual goal of the whole episode succeeded in none of the tries. Not all groups had the same number of tries per episode since a second try was only run if they did not succeed with the episode goal on the first try. That gives us a total of 82 runs of the pick behavior, where 79 were successes, see Figure 20, 12 failures. In the OR condition, there were 30 successes and 9 failures, versus 49 successes and 3 failures in the AR condition.

If the endorsement was understood correctly by the children after the first time they successfully picked up the cube, they should maybe also have understood that Cozmo first wanted them to pick up the cube in the following episodes. However, episodes 1 and 2 had only 3 failures overall conditions, all three of them in the OR condition, see Figure 8. In Figure 20 we can see that there are 9 more failures overall, 6 more in the OR condition and 3 in the AR condition, which all happened after the initial runs of the pick behavior. Consequently, it seems it wasn't always clear in the following episodes that Cozmo still wanted them to pick the cube. Although, the success rate is still very high compared to the fail-rate, so the failures might not be of consequence. Also, the AR condition still had a higher success rate and lower fail rate, mirroring the results from Episodes 1 and 2.

As for the time of the successes, the overall average was 16 seconds (M=13sec, SD=13.28sec), in the OR condition it was 18 seconds (M=14sec, SD=16.32) and 15 seconds (M=13sec, SD=11.11sec) in the AR condition. See Figure 21.

So, for the pick behavior over all the episodes, the AR condition still seems to have been better.

5 DISCUSSION

With the pilot study, we looked at if AR could effectively be used to communicate the robot's intentions when interacting with children.

Robot behaviors were tested in 6 episodes in a pilot study. These robot behaviors each had an intention that the robot tried to convey. These intentions were:

- Getting a child to pick a cube up (tested in two episodes).
- Getting a child to pick a cube up and then place it somewhere else, indicated by Cozmo.
- Getting a child to pick up a cube and hand it over to another child either directly, or placing it in front of another child and that child taking it.
- Getting a child to tap on a cube.
- Getting a child to shake a cube.

To be able to answer the research question through testing the behaviors in the only robot (OR) and augmented reality (AR) conditions, we looked mainly at how the AR condition was different from the OR condition in regards to successes and failures of finishing the task, as well as the time needed for a success.

5.1 Discussion of Results

Overall, the AR condition had more successes than the OR condition, with 75% versus 64% respectively, indicating that AR could make the robot's intentions clearer. However, the difference between the conditions was not really conclusive, especially since there was one group in the OR condition that failed all the tasks in all the episodes. If that group had been treated as an outlier, the total number of successes would have been even more similar. Nevertheless, there is still an indication that AR might be more reliable.

Relating to the specific intentions looked at, for Pick, the AR condition seems both more reliable and better time-wise. However, the success-rate for the OR conditions was also high. This could be an indication that the robot movement alone for Pick was already quite clear, and adding the AR improves the clarity, but marginally. Additionally, for all behaviors, we tried making the movement as legible as possible.

The pick behavior was also repeated as the first step for many of the other intentions. Looking at the results of all the picks, one can see that it had a high success rate over all the episodes, and AR especially had a high success rate and low fail-rate for this behavior. However, since this was the intention which was repeated the most, the children would have understood that that was the expected behavior. Especially if they understood it the first time, so the high success rate can be considered expected. Although, there are some failures that seem to have occurred after the initial successful communication of the pick intention. It is unclear what the reasons for those failures were. In some groups, the children were confused by failures for other intentions and seemed to second-guess the intention that they had previously understood correctly, and some tried exploring other ways of interacting with the cubes. Even so, the AR condition still had a slightly higher success rate and slightly lower fail-rate.

For Pick and Place, the AR condition is also slightly better. Since the AR only slightly improves the results for this intention, it could, as for Pick, be an indication that the robot movement alone was already clear and AR improves it only slightly.

For Pass Responsibility, the results between the conditions were similar to each other, especially for the first hand-over where both conditions had the same number of failures and the AR condition only had one more success than the OR condition. The task in this intention was a slightly more complicated task, pertaining to making the cube change hands between the children themselves and not only change places on the table. In both the first hand-over and the second, the OR condition had marginally better time results, although the second hand-over only had 4 successes to look at. This would indicate that the OR behavior alone might have been easier for the children to interpret than when adding AR to it. However, the AR condition was slightly better success-wise in the first hand-over, but a lot less in the second hand-over. Overall, the results lean towards OR being better for this episode, but the differences between the conditions are not large enough for that result to be conclusive. Since the action required by the children for this intention was more complicated, maybe it should have been tested separately to be able to see what worked specifically for conveying that intention, and what did not. Maybe also have a larger focus on the collaboration aspect of the intention, as well as seeing if the robot behavior alone, like in the OR condition, really was clear enough. Additionally, there were many groups where the children explicitly divided turns among themselves for when who was to try to do something with Cozmo, which might have resulted in them not handing over the cube to each other because it was a specific child's turn.

One intention that had very low results success-wise was Tap. For this intention, most groups failed to understand what Cozmo wanted them to do, in both conditions. So the robot behavior was unclear, and adding AR did not make the intention clearer. A tapping motion requires hands or fingers which Cozmo did not have. The robot motion used instead had Cozmo getting close enough to the cube with its forklift-like arm lifted for it to come down onto the cube when lowering it to "tap" the cube. However, the children in most experiment runs tended to move the cube out of the way of Cozmo, thinking that was what it wanted them to do, leading Cozmo to not being able to complete the "tapping" behavior pattern. The combination of the lack of hands and fingers making the tapping motion hard to recognize, with the fact that Cozmo often could not complete its tapping motion could have played a role in the high amount of failures for this intention. However, there were still a few groups that succeeded in the tapping task, 2 in the OR condition and 1 in the AR condition, indicating that the behavior might not have been entirely illegible, or that the children tapping the cube was a random occurrence resulting from them trying different things with the cube to get Cozmo to react.

Shaking had better results than Tap, but similar. In contrast to Tap, there were more successes than failures, but like Tap, the OR condition had one more success and fewer failures. A difference from Tap was that the AR successes were quicker. So, the results lean towards the OR condition being more reliable but the AR condition, when successful, was quicker. This could mean that the robot behavior alone was clearer and adding AR could potentially have made the intention more unclear. However, that does not explain that the AR successes were quicker, indicating that the groups of children were very different from each other for this intention. We would need to run this intention more times for more conclusive results.

The overall behavior of Cozmo was quite intelligent and responsive, since Cozmo was being controlled by a human wizard. Cozmo had intelligent functionalities that we did not use, however it is still not quite as responsive as the use of a wizard might lead one to believe. Since the technology is not on an even level with what a human wizard can do, it would be interesting to see how the children would respond to a less intelligent robot. How would the robot motion and use of AR to communicate intention be affected?

One behavior which was harder for the wizard to control perfectly was the endorsement behavior from Cozmo when the children succeeded in the task. It was in many cases slow and unclear. It was unclear in almost all experiments if the children understood when they had done it correctly and if they understood that Cozmo was endorsing them. The children being curious about Cozmo and wanting to engage with him was apparent, which reflects previous research [4]. This also means that the endorsement behavior being unclear impacted the children, or lessened the positive impact the endorsement was planned to have on them because they wanted reactions from Cozmo but the reactions they got were hard for them to interpret.

In conclusion, adding AR to help convey the social robot's intentions had a positive impact on most of the intentions. On other intentions, the robot's movements alone were already quite successful, so while adding AR could slightly improve the intent communication, a larger study would be needed to study the differences more.

5.2 About the Method

Designing an experiment for children can be complicated and can not follow the exact same model as designing experiments for adults, which we needed to account for. In hindsight, we would have benefited from a training session before the experiment for the children to get accustomed to Cozmo and some of his behaviors. However, part of the inherent design of the experiment was to record the children's first reactions, to see if the behaviors were intuitive, so we chose not to have a training session. In a more complicated game, that might not be an issue.

The experiments were videotaped and the videos were annotated to get the results. However, the children's behaviors were not always exactly what we expected and that will influence the results since the person annotating the video will need to make a judgment on how to interpret it. Since the same person annotated all the videos the inconsistencies should be minimal but they might still be there. One solution would have been to design the experiments differently so that some of the results were automated.

Another aspect of the method that is not ideal but can not be avoided is the fact that we came to the children to run tests, and needed to adapt the groups and number of children in the groups to what could be arranged. However, this is to be expected and we as researchers need to adapt. One interesting aspect of the groups of children is that they can have very different dynamics, if one child is dominant for example, or if none is, or when they autonomously give each other turns.

It was important to try to keep the child's attention and to keep the child interested, but if the child does not engage there is not much that can be done without direct interference from the experimenter, which we tried to keep to a minimum. This is what happened in the group where all episodes failed.

5.3 Resulting Guidelines

The testing of these different behaviors and if they managed to convey the wanted robot intent gives us an indication of what could be useful in more complex CRI. One such context is, for example a game using social HRI in an effort to minimize the language barriers and simultaneously promote inclusion and collaboration between newly arrived immigrant children and Swedish speaking children. However, this is outside of the scope of this thesis, so here are some guidelines for developing such a game or a similar game based on the results of this pilot experiment.

5.3.1 Pick. The Pick was the most reliable behavior and can easily be adapted to be used in a more complicated game, and AR can be used for the reliability to be even better.

5.3.2 Pass Responsibility. Getting the children to pass the cube between them would have been beneficial for a collaboration aspect. However, the results from those episodes show that it was not clear as designed in this experiment, so that would need to be redesigned or solved through some other behavior.

5.3.3 Endorsement. The endorsement is an important aspect of the interaction, but it needs to be made clearer and quicker. One solution for making it quicker would be automating more of the game and AR so that the Wizard would not need to manually activate the endorsement. Another solution could be to add modalities to the endorsement, such as sound or color.

6 CONCLUSION

The question to answer in this paper was if using AR could improve a social robot's intent communication when interacting with children. The findings show that, overall, AR did have a positive effect in most cases, making the robot's intentions clearer. However, the differences between using AR and not using AR were sometimes small, meaning that the social robot's movements alone were already quite successful. The cases where the communication of intent was the clearest was when the intention was to get the child to pick a cube. We recommend that behavior be used in future developments. More complicated actions like passing the cube between children did not have clear enough results for it to work as designed in this experiment, so this would need to be redesigned or not used at all.

Endorsing the children when they did what was expected when they understood the social robot's intentions, needed to be clearer. It is important when working with children to keep them engaged. They are also often inclined to be curious about the robot and interaction, which can be taken advantage of. We recommend the endorsement to be quicker and clearer to meet the children's curiosity and engagement better.

In conclusion, this paper looked at specific behaviors of a social robot, one by one, to see if AR could enhance them and make the robot's intentions clearer. The next step would be to design a more complex interaction, like a game with the behaviors and AR that worked or the suggested developments, combining them and testing the potential of AR in a non-verbal context such as promoting inclusion and collaboration between newly arrived immigrant children and Swedish speaking children for example.

ACKNOWLEDGMENTS

I want to thank Sarah Gillet and Gaspar Isaac for their help in running the experiment, the children who participated, as well as the support from my supervisor Iolanda Leite.

REFERENCES

[1] Rasmus S Andersen, Ole Madsen, Thomas B Moeslund, and Heni Ben Amor. 2016. Projecting robot intentions into human environments. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 294–301.

[2] Anki. [n. d.]. Meet Cozmo. https://anki.com/en-us/cozmo.html, Last accessed on 2019-06-10.

[3] Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, Valentin Enescu, et al. 2013. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction* 1, 2 (2013), 33–53.

[4] Tanya N Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, and Sarah Nugent. 2011. Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies* 69, 7-8 (2011), 539–550.

[5] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J Lilienthal. 2015. That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *2015 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.

[6] Nuno Costa and Artur Arsenio. 2015. Augmented reality behind the wheel-human interactive assistance by mobile robots. In *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*. IEEE, 63–69.

[7] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 301–308.

[8] Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 334–343.

[9] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[10] Shayegan Omidshafiei, Ali-Akbar Agha-Mohammadi, Yu Fan Chen, Nazim Kemal Üre, Jonathan P How, John L Vian, and Rajeev Surati. 2015. Mar-cps: Measurable augmented reality for prototyping cyber-physical systems. In *AIAA Infotech@ Aerospace*. 0643.

[11] Emanuele Ruffaldi, Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli. 2016. Third point of view augmented reality for robot intentions visualization. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 471–478.

[12] Suleman Shahid, Emiel Krahmer, and Marc Swerts. 2014. Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human Behavior* 40 (2014), 86–100.

[13] Jinglin Shen, Jingfu Jin, and Nicholas Gans. 2013. A multi-view camera-projector system for object detection and robot-human feedback. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 3382–3388.

[14] Elena Sibirtseva, Dimosthenis Kontogiorgos, Olov Nykvist, Hakan Karaoguz, Iolanda Leite, Joakim Gustafson, and Danica Kragic. 2018. A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 43–50.

[15] Sichao Song and Seiji Yamada. 2017. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2–11.

[16] Fumihide Tanaka, Javier R Movellan, Bret Fortenberry, and Kazuki Aisaka. 2006. Daily HRI evaluation at a classroom environment: reports from dance interaction experiments. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 3–9.
[17] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 316–324.
[18] Tom Williams. 2018. A framework for robot-generated mixed-reality deixis. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

TRITA -EECS-EX-2019:244