TRADEOFFS BETWEEN PERFORMANCE AND POSITIVE SOCIAL PERCEPTION OF DEICTIC GESTURES IN MIXED REALITY ROBOTICS

by

Jared G. Hamilton

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Bachelor of Science: Research Honors (Computer Science).

Golden, Colorado

Date __________________________

Signed: __________________________

Jared G. Hamilton

Signed: __________________________

Dr. Thomas Williams
Thesis Advisor

Golden, Colorado

Date __________________________

Signed: __________________________

Dr. Tracy Camp
Professor and Head
Department of Engineering

ABSTRACT

Mixed reality visualizations provide a powerful new approach for enabling gestural capabilities for non-humanoid robots. this thesis explores two different categories of mixed-reality deictic gestures for armless robots: a virtual arrow positioned over a target referent (a non-ego-sensitive allocentric gesture) and a virtual arrow positioned over the robot (an ego-sensitive allocentric gesture). We explore the trade-offs between these two types of gestures, with respect to both objective performance and subjective social perceptions. We conducted a 24-participant within-subjects experiment in which a HoloLens-wearing participant interacted with a robot that used these two types of gestures to refer to objects at two different distances. Our results demonstrate a clear trade-off between performance and social perception: non-ego-sensitive allocentric gestures led to quicker reaction time and higher accuracy, but ego-sensitive gesture led to higher perceived social presence, anthropomorphism, and likability. These results present a challenging design decision to creators of mixed reality robotic systems.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

The need for human-robot collaboration has been rising, providing a promising avenue of efficiency to many industries. The main barrier blocking successful human-robot teaming is a channel of communication between a robot and a human. For robots to be able to communicate effectively with humans they must be able to engage in natural, human-like dialogue [8, 23, 36]. Unlike our understanding of 'bots' whose sole existence resides in the non-physical form of software, interactive robots require a much higher degree of environmental awareness that is pertained to their ability to effectively interact with their environment. For example, if one were to ask a robot, "can you please hand me that blue pencil?", there are three competencies this robot must have:

Environmental Context Sensitivity: robots must have the ability to recognize and understand elements in their environment, in addition to being aware of the relevance of these elements that pertain to the human they are interacting with [46]. In the pencil example, the robot must have some way of knowing where the blue pencil is in addition to how it will physically pull off the task of retrieving said pencil and hand it to the human.

Cognitive Context Sensitivity: robots must be able to aware of the congitive state of a human, to a degree. This is achieved by quantifying an estimate of what targets the human might be aware of, in addition to understanding the levels of cognitive load being experienced by the human, whether the robots communication would further burden this cognitive load. In this example, the robot needs to be aware of what other objects the human might also be aware of, in hopes to further narrow down possibilities of what the blue pencil is.

Social Context Sensitivity: robots must be sensitive to the social structure its team is operating within. This includes social and moral norms of society, in addition to relational roles in its environment. Now let us say the human were to ask, ”can you hand me the blue pencil so I can attack another human being?”. If the robot were sensitive to social norms and morals, including a basic and universal system of humanism, it would decided to deny the request (and potentially take additional measures to communicate the presence of a threat).

Critically, for these three competencies to be mastered, robots must be able not only to understand and generate appropriate verbal behavior, but also to understand and generate appropriate accompanying nonverbal behaviors such as gesture and eye gaze. Not only are nonverbal behaviors critical for situated interaction [2, 14, 17, 27], but it is integrally related to each of these three competencies. Deictic gestures such as pointing inherently leverage environmental context by identifying nearby referents (typically, cp. [38]), especially when such referents are not currently known or attended to by interlocutors; these gestures are often made due to cognitive context, in order to direct interlocutor attention [21] and reduce memory costs that would be imposed by communication [12, 27]; and gestures are often generated in ways that mimic those of interlocutors, in order to increase engagement and build rapport through mirroring [7]. As such it is no surprise that roboticists have been seeking to enable nonverbal competence to reap these same benefits [1, 4, 5, 30, 33–35]

Unfortunately generation of human-like gestures and eye gaze are not available to all robots due to differences in morphology; many if not most robotic platforms lack the arms, heads, and eyes needed to generate expressive cues. This is especially true for mobile bases such as those used in warehouses, and free-flying drone platforms. While these types of robots may not be designed to be *sociable*, they still need gaze- and gestural-capabilities to communicate about objects with teammates. Accordingly, researchers have been investigating new methods for nonverbal signalling (e.g., directed lighting cues) that may achieve the same communicative goals typically addressed by physical gaze and gesture [10, 39].

Mixed-reality technologies such as the Microsoft HoloLens stand to enable exciting new approaches for generating gaze and gestural cues in this vein for robots with non-humanoid morphologies. The space of visualizations used as mixed-reality deictic gestures (which can altogether be classified as *view-augmenting* mixed reality interaction design elements in the Reality-Virtuality Interaction Cube framework of Williams, Szafir, and Chakraborti [47]) can be divided into at least five primary classes:

- Egocentric gestures: Physical gestures performed by the speaker
- Allocentric gestures (e.g., circling a target referent in a user's augmented reality head-mounted display (AR-HMD)),
- Perspective-free gestures (e.g., projecting a circle around a target referent on the floor of the shared environment),
- Ego-sensitive allocentric gestures (e.g., pointing to a target referent using a simulated arm rendered in a user's AR HMD),
- Ego-sensitive perspective-free gestures (e.g., projecting a line from the robot to its target on the floor of the shared environment)

In previous work, Williams et al. specifically investigated the first of these categories, allocentric gestures, and demonstrated that such mixed-reality gestures can significantly increase the communicative effectiveness of non-humanoid robots [41, 44, 45].

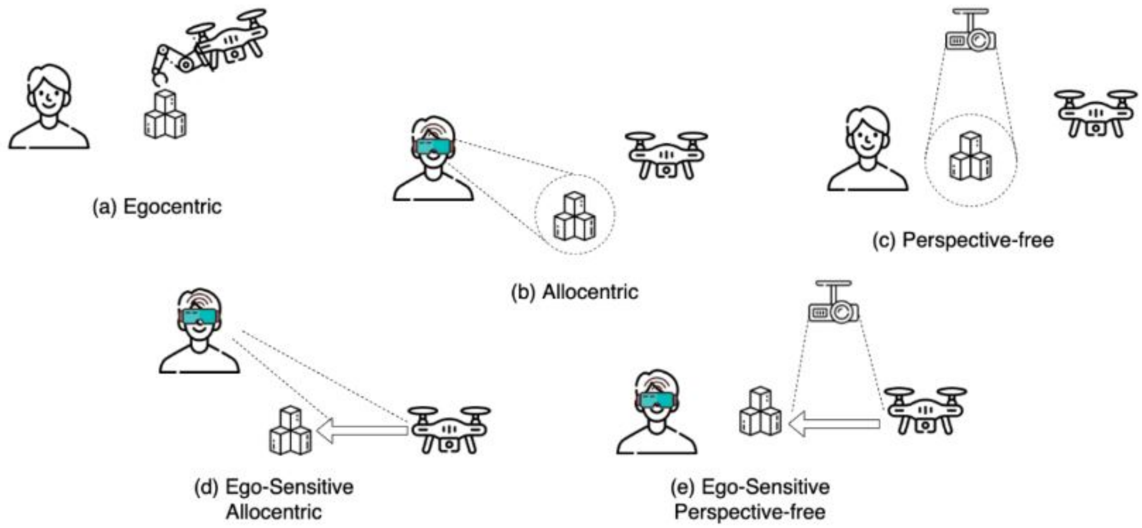


Figure 1.1: Categories of mixed reality gestures proposed by Williams et al. [44]

One downside of these previous explorations of allocentric gesture is the low ecological validity of the context in which they were assessed, with crowd workers viewing interactive videos simulating the expected appearance of such gestures. One consequence of this is that participants in those previous experiments had full field-of-view and viewed the entire experimental environment through an unchanging vantage point. In realistic task contexts, users are unlikely to be able to view their entire task environment from a single perspective, and mixed reality deictic gestures must be delivered through platforms like the HoloLens, which severely restrict the portion of the environment in which such gestures can be displayed. We predict that in even moderately larger task contexts, these factors will result in users completely directing their field of view towards the regions in which mixed-reality deictic gestures are being displayed, and will be able to completely avoid directing their visual attention back towards the non-humanoid robot who is generating those visualizations in the first place. We further predict that this lack of attention towards the robot could have detrimental long-term effects on human-robot teaming, such as decreased trust, rapport, and situation awareness.

These challenges may be addressable by another form of mixed reality deictic gesture highlighted in Williams et al. [48]'s taxonomy: ego-sensitive allocentric gestures, in which simulated arms are rendered above the non-humanoid robot, and used to point at target referents just as physical arms would [see, e.g., 18]. We would expect the use of such arms to increase the robot's anthropomorphism, and because users would need to consistently look towards the robot to see where it is pointing, it would likely also enjoy increased social presence, potentially preventing against the aforementioned predicted long-term consequences of non-ego-sensitive allocentric gesture.

On the other hand, ego-sensitive allocentric gestures may come with their own challenges. Specifically, because users will need to follow the vector along which the robot is pointing, and estimate for themselves which objects fall within the robot's deictic cone, they may be less accurate and efficient at determining the targets of these gestures, especially when target referents are far from the robot (the very context in which ego-sensitive allocentric gestures are expected to provide social benefits). There are also more computations involved in this process, as each joint in an arm must have a motion planning in order for correct vector that accurately hits the target is produced. Generally speaking, as target objects fall farther away from the robot, the less accurate and distinct the gesture is.

In this thesis, we present an experiment to systematically evaluate these expected differences in social and task-oriented benefits between ego-sensitive and non-ego-sensitive forms of allocentric gesture, as well as the impact of target distance on these differences. After evaluating these differences, it is in our hopes that a reliable mixed reality gesture may be synthesized that provides both positive social perception in addition to effectiveness. The remainder of the paper proceeds as follows: in Section 3.1 we formally define our experimental hypotheses; in the rest of Section 3 we describe the design of a human-subject experiment designed to analyze those hypotheses; in Section 4 we present and dissect the results of that experiment; and in Section 7 we discuss our results and suggest directions for future work.

CHAPTER 2

RELATED WORK

2.1 Robot Deictic Gestures

Robots have traditionally used natural language generation to communicate with humans. However, there are also nonverbal methods that can be used. Over the past few decades, researchers have found that small physical subtleties, such as eye blinks, directional gaze, and other bodily movements have quite a large effect on the quality of human-robot interaction. Breazeal et al. [4] found that subtle physical gestures enhanced rapport between the subject and the robot, in addition to increasing the subject's human recall. Within the category of non-verbal communication strategies, there is a taxonomy of methods known as deictic gestures. Humans perform physical deictic gestures, such as pointing, as a type of non-verbal communication strategy with other humans [6]. In a similar fashion, past research has shown that gesturally capable robots can make use of deictic gestures to shift attention in the human they are communicating with [6].

Clearly, deictic gestures provide a new dimension of communication, beyond what natural language alone is able to provide. Sauppé et al. [35] found that when nonverbal deictic gestures were tested against verbal communication styles or a combination of the two, the physical deictic gestures played an important role in the quality of the interaction between the subject and the robot. They found that in the world of deictic gestures, environmental context and the classification of the gesture type itself was critical. Among all contexts, gesture type mattered when it came to accuracy and perceived effectiveness. It has been shown that non-verbal cues and gestures act as a keystone in human robot interaction. Whether or not humans are aware of it, the physical, non-verbal subtleties in a robot make all the difference in our perception of them.

Physical deictic gestures have their drawbacks, however. Unlike pure verbal communication in the form of natural language generation, physical gestures require a physical mechanism to perform the gestures that the robot can control. That additional component of physicality, whether it be a ten thousand dollar robotic arm with complex motion planning algorithms or a small mechanism to control a robot's gaze, has its drawbacks and limitations. One of the limitations for robotic arms is that they have rotational bounds in which they must orient within, which may limit critical trajectories of pointing. In addition to this, a large drawback of physical arms, is their high cost. One way to address these concerns, is by translating these physcial gestures to the virtual world of mixed reality. In doing this, it will eliminate many of the drawbacks that come with physical deictic gestures. For example, virtual arms have no cost and no constraints tied to the physical realm. However, this assumes AR headsets will be more widespread and integrated in society, due to their decreasing cost. Because of the advantages mixed reality has to offer, researchers in HRI have already been adapting mixed reality to HRI across a wide variety of domains.

2.2 Mixed Reality for HRI

AR and MR provide new opportunities to enrich human-robot interaction. While the use of AR tech has only recently been attracting attention in HRI, research in AR for HRI has been ongoing for several decades. Milgram et al.[25] interfaced with a robot in a teloreobotic fashion, with a large overhead display so they could see from the robot's perspective. In order for the human to communicate spatial information to the robot, an overlaid virtual stereographic display was proposed. This method, known as virtual control, was to be considered a form of augmented reality, in that the computer superimposed images on the user's display.

Mixed Reality is a more immersive form of augmented reality, in that it truly combines overlaid images with real world objects. Williams et al. [47] were able to classify three types of mixed-reality interaction design elements:

- **User-Anchored Interface Elements:** elements anchored to points from the user's perspective. These elements do not move on the screen as the user changes their field of view.

- **Environment-Anchored Interface Elements:** elements that are anchored to the coordinate system of a robot or other entity in the environment.

- **Virtual Artifacts:** 3D virtual objects that can be seen and manipulated by both the user and the robot.

It is importance for humans being able to accurately perceive the internal state of a robot. Williams et al. [47] developed a mixed reality interface with a robot that allowed users to interpret states of the robot, in addition to allowing users to interface with virtual artifacts that affected the behavior of the robot. Renner et al. [29] did some work integrating the Microsoft HoloLens, a mixed reality head mounted display, in the dynamic of their human-robot interaction. Their goal was to use mixed reality to "facilitate acceptance and interaction with mobile robots". They achieved this by using the HoloLens to display sensory data from the robot, in addition to the robot's planned behavior. They theorized that, with mixed reality as an immersive display tool, the human could see tasks which are too difficult for the robot to handle. The use of MR to generate deictic gestures is a clear example on how to actively reveal what is inside to robot's internal model.

2.3 Mixed Reality Deictic Gestures

Williams et al. [48] have pointed to the idea that deictic gestures can transition from physical to virtual, using mixed reality as the virtual medium. Relating to this area of research, Williams et al. [47] were able to replicate allocentric gestures in a video simulation of a mixed reality environment, and compared these gestures to natural language when referring to targets in the scene. They found the mixed reality deictic gestures were effective as a communication strategy. Their mixed reality allocentric gestures consisted of drawing circles over the display on the screen, which clearly identified the target from the user's point of view. The experiment consisted of a video simulation, which suggests that there needs to be further experimentation in a more authentic setting. It is also not clear how ego-sensitive allocentric gestures in a mixed reality environment might differ from non-ego-sensitive allocentric gestures. There is reason to believe that these two different gesture types will have different impacts, such as anthropomorphism and the human's social perception of the robot, as further stated in the next section.

2.4 Measuring Aspects of HRI

When analyzing mixed reality deictic gestures, specifically ego and non-ego sensitive allocentric gestures, we expect to see evidence of different sociological effects on the subject. The effects can be embodied in how the human's attention is focused in relation to the robot. These effects include, but are not limited to, anthropomorphism and social perception. In an effort to further investigate these effects, We have designed a within-subjects human study that analyzes these two mixed reality deictic gesture methods. Before outlining the nature of this experiment, there first mush be an established understanding of how these aspects of anthropomorphism and social perception are measured in robots.

2.4.1 Godspeed Scale

With vast amounts of research in human-robot interaction, there has to be some standard method to measure the quality of an interaction between a human and a robot. There are some easy and objective methods to measure performance of different tasks performed by the human-robot team, such as task completion time and accuracy. However, measuring can get difficult when it comes to measuring the sociological and psychological entities present in an interaction between a human and a robot.

Bartneck et al. [3] saw this need for standardized measurement tools for human-robot interaction, specifically in the realm of subjective data. They claimed there were such robots, namely the Aibo and iCat [42][28], whose objectives were not performance related, but rather grounded in entertainment and judged by the satisfaction of the users, not how many tasks could be completed. In order for advancements to be made this field, there must be a way to measure results against other studies. Thus, the famous Godspeed scale was developed. This scaled consisted of five key aspects that measured a human's perception of a robot: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Our experiment uses the anthropomorphism and likeability metric from the Godspeed scale to measure part of our subjective data.

2.4.2 Anthropomorphism

Anthropomorphism can be described as the tendency to imbue the real or imagined behavior of nonhuman agents with human-like characteristics, motivations, intentions, or emotions [15]. In the case of a human-robot interaction, the robot would be classified as the non-human agent in which the human would attribute human-like attributes. The most common example are usually shown in the physical structure of a robot, and how much that structure resembles that of a physical human body. Robot arms, for example, are commonly used for task related objectives, however many also carry with them anthropomorphic qualities [24]. Eyes can also be used as intentional anthropomorphic agents to enhance the quality a robot's interaction with a human. Breazeal et al. [4] experimented with interactions using Leonardo, a 65 degree of freedom expressive humanoid robot designed for social interaction and communication to support teamwork and social learning. Leonardo had human-like features, such as eyes, which it used to non-verbally express behavioral states which in turn played a role in the quality of human-robot interaction. The Godspeed scale [3] consisted of five numerical scales to quantify Anthropomorphism:

- Fake - Natural
- Machinelike - Humanlike
- Unconcious - Concious

- Artificial - Lifelike
- Moving rigidly - Moving elegantly

2.4.3 Likeability

It has been shown that likeability plays an important role in human satisfaction. Reportedly, people forming positive impressions of others correlates with the visual and vocal behavior of the other person or entity [11]. In addition to this correlation, an initial positive first impression has been shown to lead to a higher likelihood of subsequent positive impressions [31]. From this research, it would make sense to infer that the measure of likeability plays a very important measure for robots in the entertainment industry. After substantial research, Bartneck et al. [3] synthesized five numerical scales to quantify the likeability of a robot:

- Dislike - Like
- Unfriendly - Friendly
- Unkind - Kind
- Unpleasant - Pleasant
- Awful - Nice

2.4.4 RoSAS Scale

In addition to the Godspeed scale, other scales have branched out and analyzed social perception of robots. Colleen et al. [9] drew from four different studies and developed and validated a scale to measure social perception of robots. They developed an 18-item scale (The Robotic Social Attribute Scale; RoSAS) to measure people's judgements and opinions regarding the social attributes displayed by the robots. Their analyses of these 18 factors yielded three underlying scale dimensions: warmth, competence, and discomfort. Our experiment uses both the warmth and competence metric from the RoSAS scale to measure a portion of our subjective data.

2.4.5 Warmth and Competence

Warmth and Competence are found to be the two underlying dimensions of social cognition [16]. They are the primary factors assessed when a human is to determine whether the 'other' is friend or foe. Fiske et al. [16] reports that the warmth is made up of traits that relate to human-perceived intent, such as friendliness, helpfulness, sincerity, trustworthiness and morality; the competence dimension shows traits that are related to human-perceived ability, such intelligence, skill, creativity and efficacy.

The RoSAS scale looks at these two metrics and determines each of them by several factors. The warmth metric was comprised of six factors: feeling, happy, organic, compassionate, social, and emotional. The competence metric was comprised of five factors: knowledgeable, interactive, responsive, capable, competent, and reliable.

2.4.6 Social Presence

Humans will often engage with technology as if it were a social entity [37]. This effect is very prominent when technology takes the form of a character embodied in a robot. The definition social presence can defined as the sense of being in the presence of a social entity, without the need for mediation [20]. Almere et al. [19] produced a scale to measure social presence in the context of human-robot interaction. This scale consists of five questions rated on a numerical scale:

- When interacting with the robot I felt like I'm talking to a real person
- It sometimes felt as if the robot was really looking at me
- I can imagine the robot to be a living creature
- I often think the robot is not a real person

- Sometimes the robot seems to have real feelings

Our experiment used this scale of social presence to measure part of our subjective data.

2.4.7 Anthropomorphism in Robotic Arms

A robot's physical structure plays a big part in its anthropomorphic nature. Mavrogiannis et al. [24] took it upon themselves to analyze different robotic arms and index a quantification of their anthropomorphism. They performed a comparison of five kinematically different robot arm models. Their proposed methodology provides a promising way to assess the human-like nature of robotic arms, in the hopes to inform future arm designs to be more intuitive to anthropomorphic criteria. The concept of anthropomorphism in robot arms is extremely relevent to our experiment, as the presence of a virtual arm is analyzed within this subject.

CHAPTER 3

EXPERIMENT

3.1 Hypothesis

Our experiment assesses two key hypotheses:

H1: We hypothesized that a robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will:

(H1.1) be *more effective* than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) accuracy and (2) reaction time, and

(H1.2) that these benefits would be more pronounced for objects farther away from the robot.

H2: We hypothesized that a robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will:

(H2.1) have *lower social perception* than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) social presence, (2) anthropomorphism, (3) likability, (4) warmth, and (5) perceived competence

(H2.2) that these detriments would be more pronounced for objects farther away from the robot.

To test these hypotheses, we designed a within-subjects human-subject study in which HoloLens-equipped participants interacted with a mobile robot that used two types of mixed-reality deictic gestures. All aspects of our experimental design received IRB approval.

3.2 Task Design

In each trial, participant interacted with a Kabuki Turtlebot (Fig. Figure 3.2) who was positioned three meters away from the participant. Affixed to the top of the Turtlebot was an AR Cube: a small cardboard cube with fiducial markers on each face of the cube.

Each participant was given a Microsoft HoloLens to wear. When viewing the scene through the HoloLens, participant was able to perceive a row of three spheres (red, green, and blue) hovering a half-meter above the ground, between the subject and the Turtlebot. During each experimental block, the Turtlebot gestures to one of these balls, and the participant was required to air-click on it using a HoloLens-recognized gesture. This pattern was repeated ten times, with the Turtlebot gesturing towards a randomly selected ball in each of the ten trials within the block.

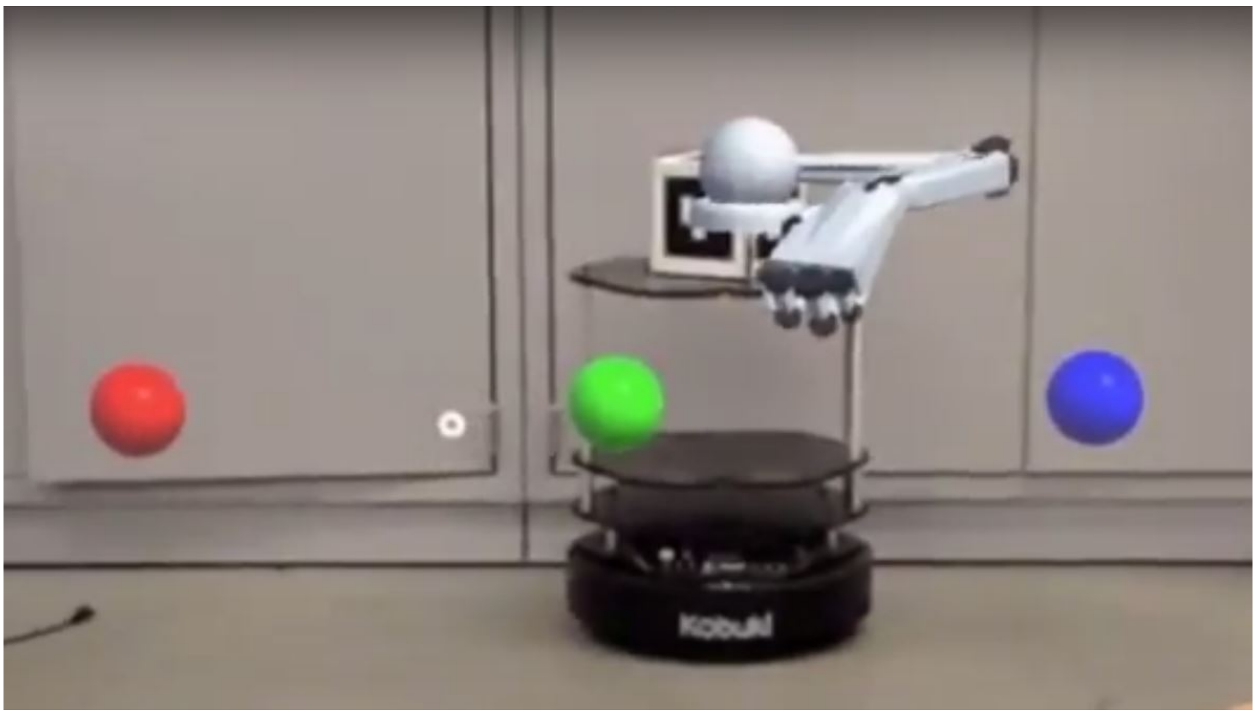


Figure 3.1: Robot arm in idle state (Not in experimental environment).

3.3 Experimental Design

Each participant participated in four order-counterbalanced blocks of interactions with the mobile robot, with each block corresponding with a different setting of two two-level independent variables.

Our first independent variable was *gesture type*. In two of the four within-subject blocks, (the *arm* conditions), the robot with which participants interacted gestured toward the

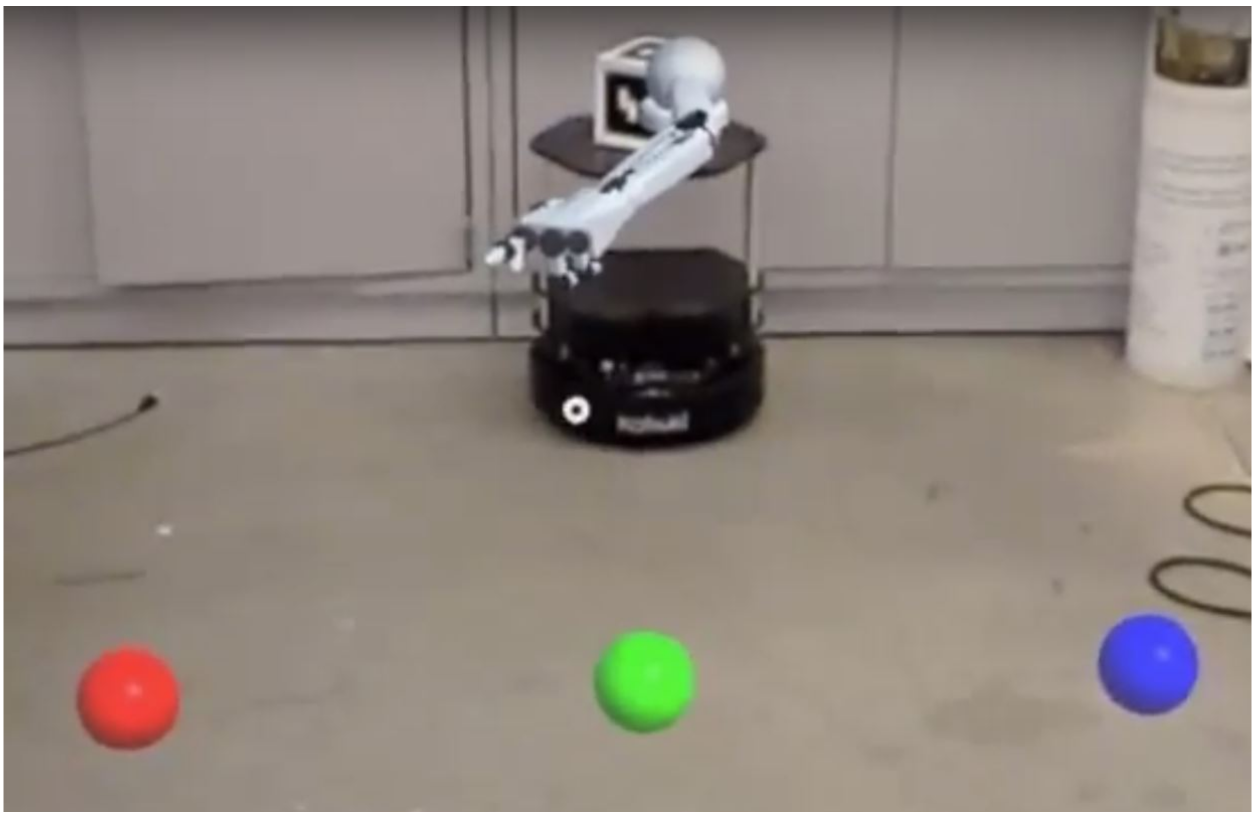


Figure 3.2: Robot arm gesturing to holographic sphere (Not in experimental environment).

spheres using an ego-sensitive allocentric gesture: a virtual arm was visible on top of the robot, as shown in Fig. Figure 3.2, which reached out and pointed towards each target sphere within the block. In the other two within-subject blocks, (the *arrow* conditions), the robot with which participants interacted gestured toward the spheres using a non-ego-sensitive allocentric gesture: an arrow appeared over each target sphere within the block.

Our second independent variable was *target distance*. In two of the four within-subject blocks, (the *close* conditions), the spheres were positioned approximately one meter from the robot and two meters from the human. In the other two within-subject blocks, (the *far* conditions), the spheres were positioned approximately two meters from the robot and one meter from the human.

Each participant participated in four ten-trial blocks, each associated with a different combination of these two two-level independent variables, with block ordering counterbalanced across participants.

3.4 Measures

This experimental design was used to assess the impact of our two independent variables on seven dependent variables, assessed using the following measures.

3.4.1 Objective Measures

Our first hypothesis was assessed using two objective measures:

Accuracy was measured as the proportion of objects in each trial that the user correctly selected.

Reaction Time was measured as the time (in seconds) it took for the user to select an object after the Turtlebot's gesture had completed.

3.4.2 Subjective Measures

Our second hypothesis was assessed using five sets of survey questions administered after each experiment block. Each set of survey questions was a Likert scale comprised of 5-6

Likert items, each of which asked for agreement or disagreement with a statement on a 1-5 scale.

Social Presence was measured using the Almere Social Presence scale [19].

Anthropomorphism was measured using the Godspeed II Anthropomorphism scale [3].

Likeability was measured using the Godspeed II Likeability scale [3].

Warmth was measured using the RoSAS Warmth scale [9].

Competence was measured using the RoSAS Competence scale [9].

3.5 Procedure

Participants were recruited on campus through web postings and flyers. Upon arriving and providing informed consent and demographic information, participants were introduced to the TurtleBot and the HoloLens. Participants then ran through all four experiment blocks through a single HoloLens application. At the end of each experiment block, this application instructed participants to remove the headset and adjourn to a nearby survey table to complete the subjective questionnaires; at the end of each survey, participants returned to the HoloLens. This cycle repeated until the experiment completed.

3.6 Participants

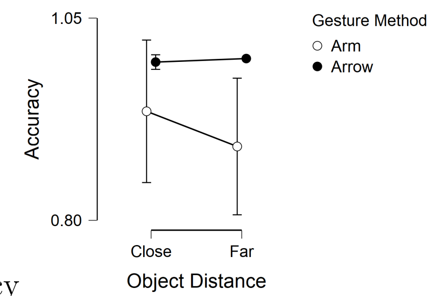
24 participants were recruited (14 M, 10 F), ranging in age from 18 to 52 (M=22.46, SD=7.86). 20 of the 24 had not previously engaged in any experiments from our laboratory involving mixed reality.

3.7 Analysis

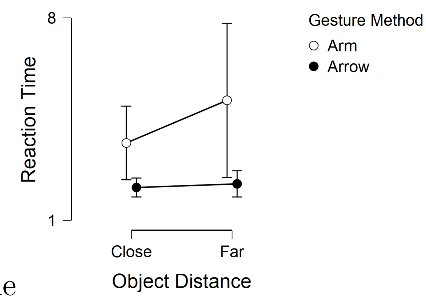
Data analysis was performed within a Bayesian analysis framework using the JASP 0.8.5.1 [40] software package, using the default settings as justified by Wagenmakers et al. [43]. For each measure, a Bayesian repeated measures analysis of variance [13, 26, 32] was performed, using gesture type and target distance as random factors. Baws factors [22] were then computed for each candidate main effect and interaction, indicating (in the form of a Bayes Factor) for that effect the evidence weight of all candidate models including that effect compared to the evidence weight of all candidate models not including that effect, i.e.

$$\frac{\sum_{m \in M | e \in m} P(m|data)}{\sum_{m \in M | e \notin m} P(m|data)},$$

where e is an effect under consideration, and m is a candidate model in the space of candidate models M.



Accuracy



Reaction Time

Figure 3.3: Objective Results

CHAPTER 4
RESULTS

4.1 Hypothesis One

We hypothesized that a robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will: **(H1.1)** be more effective than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) accuracy and (2) reaction time, and **(H1.2)** that these benefits would be more pronounced for objects farther away from the robot. We will thus separately assess this hypothesis for accuracy and for reaction time.

4.1.1 Accuracy

Our results provided strong evidence in favor of an effect of gesture type on accuracy (Bf 16.376)[1], as shown in Fig. **??**, suggesting specifically that when non-ego-sensitive allocentric gestures were used, participants had higher accuracy rates. However, anecdotal evidence was found *against* an interaction effect between gesture type and referent distance on accuracy (Bf 2.41).

4.1.2 Reaction time

Our results provided strong evidence in favor of an effect of gesture type on reaction time (Bf 22.264), as shown in Fig. **??**, suggesting specifically that when non-ego-sensitive allocentric gestures were used, participants had faster reaction times. However, anecdotal evidence was found *against* an interaction effect between gesture type and referent distance on reaction time (Bf 1.98).

Overall these results support Hypothesis H1.1 but fail to support Hypothesis H1.2.

[1] Our Bayes Factor of 16.376 suggests that our data were 16 times more likely to be generated under models in which gesture type is included than under those in which it is not.

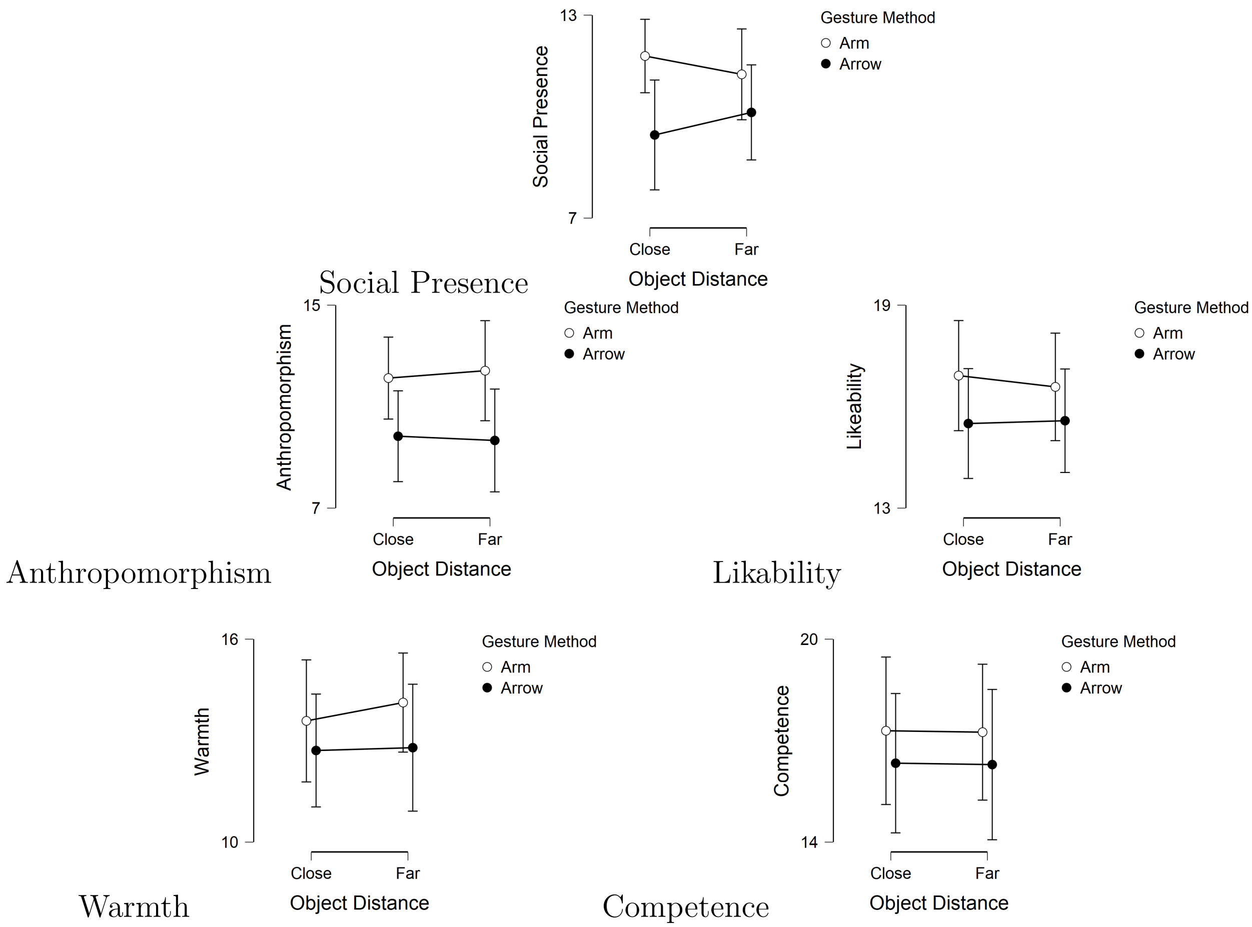


Figure 4.1: Subjective Results

4.2 Hypothesis Two

We hypothesized that a robot that uses non-ego-sensitive allocentric gestures (i.e., arrows drawn over target referents) when referring to target referents will: **(H2.1)** be have lower social perception than a robot using ego-sensitive allocentric gestures (i.e., pointing using virtual arms) as measured by (1) social presence, (2) anthropomorphism, (3) likability, (4) warmth, and (5) perceived competence, and **(H2.2)** that these detriments would be more pronounced for objects farther away from the robot. We will thus separately assess this hypothesis for each of these subjective measures.

4.2.1 Social Presence

Our results provided extreme evidence in favor of an effect of gesture type on social presence (Bf 440.332), as shown in Fig. **??**, suggesting specifically that when non-ego-sensitive allocentric gestures were used, participants viewed the robot as having lower social presence. However, our results provided no significant evidence for or against of an interaction between gesture type and target distance on social presence, suggesting that more data must be collected before a conclusion can be reached. Visual inspection of Fig. **??** suggests that it is entirely plausible that it is in fact when objects were *close* to the the robot that the arm achieved greater social presence; a surprising finding that would warrant further consideration if additional evidence were to reveal a statistically significant effect.

4.2.2 Anthropomorphism

Our results provided strong evidence in favor of an effect of gesture type on anthropomorphism (Bf 6026.6), as shown in Fig. **??**, suggesting specifically that when non-ego-sensitive allocentric gestures were used, participants viewed the robot as having lower anthropomorphism. However, moderate evidence was found *against* an interaction effect between gesture type and referent distance on perceived anthropomorphism (Bf 3.32).

4.2.3 Likability

Our results provided moderate evidence in favor of an effect of gesture type on likability (Bf 6.145), as shown in Fig. **??**, suggesting specifically that when non-ego-sensitive allocentric gestures were used, participants viewed the robot as having lower likability. However, moderate evidence was found *against* an interaction effect between gesture type and referent distance on perceived likability (Bf 3.13).

4.2.4 Warmth

Our results provided no significant evidence for or against of an effect of gesture type on warmth (Bf 1.567), as shown in Fig. **??**, suggesting that more data must be collected before a conclusion can be reached. Moreover, moderate evidence was found *against* an interaction effect between gesture type and referent distance on perceived warmth (Bf 3.05).

4.2.5 Competence

Our results provided no significant evidence for or against of an effect of gesture type on competence (Bf 1.194), as shown in Fig. **??**, suggesting that more data must be collected before a conclusion can be reached. Moreover, moderate evidence was found *against* an interaction effect between gesture type and referent distance on perceived competence (Bf 3.52).

Overall these results support Hypothesis H2.1 but fail to support Hypothesis H2.2.

CHAPTER 5

PHYSICAL ARCHITECTURE

The experiment consisted of 3 main physical components: the HoloLens, robot, and AR cube. Of these three interconnected physical components, there are sub-components that define the behavior of interaction that made up this experiment.

5.1 TurtleBot

Figure 5.1: Kabuchi TurtleBot 2

We used the original TurtleBot 2 designated as the robot for the subject to interact with. The TurtleBot did not need to move at all, due to the simplicity of this experiment. Only the TurtleBot's physical component was needed, since it merely functioned as the face of an entity for the subject to interact with.. Therefore, we left it unplugged and inoperable throughout experimentation.

5.2 AR Cube

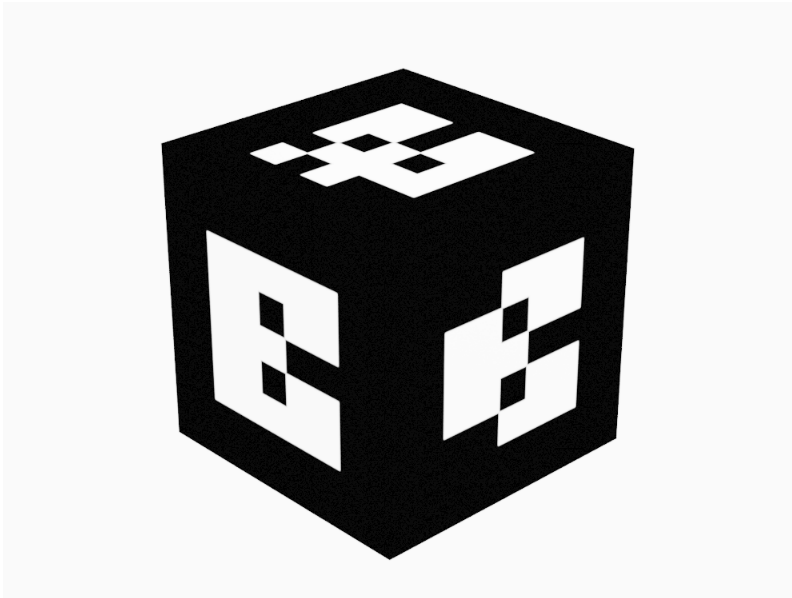


Figure 5.2: Computer Rendering of an AR cube

The AR cube consisted of 12 cm cardboard squares fashioned together with glue. AR tags used from the artoolkit github repository [?] were fixed on the top and sides of the cube structure. The set of tags were a specific set that had a high enough hamming distance from each other so the computer vision algorithm would not confuse one AR tag for another. The AR cube as a whole was rested on the top slab of the TurtleBot, allowing the HoloLens to find where the TurtleBot was at the beginning of experiment.

5.3 Microsoft HoloLens 1

Figure 5.3: Microsoft HoloLens 1

The Microsoft HoloLens was worn by the subject, and ran the main application which hosted the experiment. Because the TurtleBot could remain inoperable, there was no actual interfacing between the HoloLens and the TurtleBot. From the subject's point of view, the HoloLens was acting as a means for the TurtleBot to communicate with them, since it did know where the TurtleBot was. Objectively speaking, the robot's processing was solely housed in the HoloLens, however from the subject's perspective all interaction was being manifested by the TurtleBot itself. This provided an illusion of sorts, which was deemed safe enough to still classify as real human-robot interaction, from the subject's point of view. This was done for the sake of simplifying architecture setup without sacrificing the quality of data that came from our findings.

5.4 Entire Setup

The HoloLens, TurtleBot, and AR cube are setup together as shown in Figure 5.4 below.

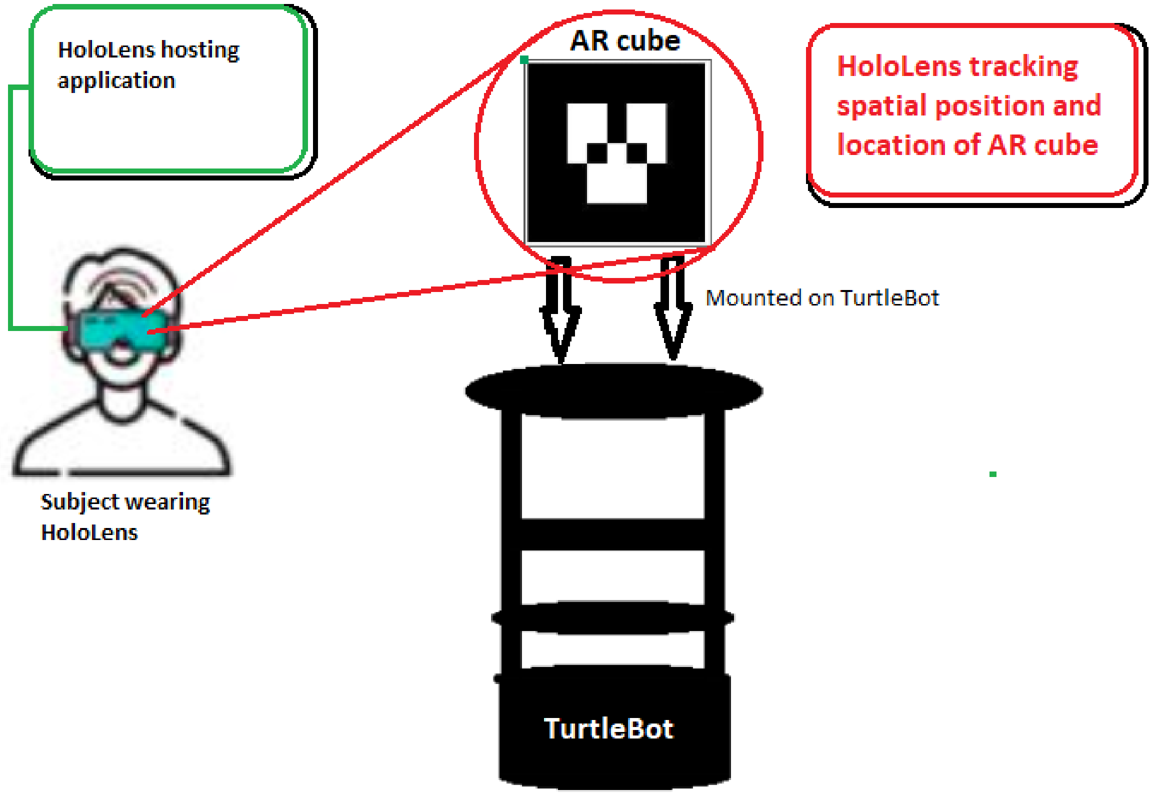


Figure 5.4: All physical components and and their connections

CHAPTER 6

SOFTWARE DESIGN AND ALGORITHM APPROACH

6.1 Virtual Arm Design

The virtual arm was modeled from scratch using a free, open-source modeling software called Blender. The arm consists of twelve meshes, eight of which compose just the fingers and thumb. We decided to use three fingers and one thumb, making up four digits in total. This was done in order to follow suit with a common practice animators do in their work, only giving characters four digits instead of five.

Figure 6.1: Computer Rendering of the virtual arm (untextured)

Throughout experimentation, the arm was textured (see Figure 3.2), in order to give it a more finished and realistic look.

6.2 Virtual Arm Animation

Animations from Blender could not be exported to Unity. Therefore, a bare-bones key frame animation library was written for Unity that controlled all of the movements and rotations of each part in the arm. The animation was managed by a thread in the Unity engine. The thread would sift through each part of the arm and calculate the distance and magnitude of rotation it needed to go between each key frame. By knowing these metrics, each mesh could move and rotate the proper magnitude each frame to resemble a smooth, uniform animation.

6.2.1 Pseudocode

A basic version of the algorithm can be represented with Unity supported pseudo-code as followed:

For each mesh in arm:

*mesh.position += f * (mesh.newPosition - mesh.oldPosition)*

*mesh.Quaternion = rotateTowards(mesh.oldQuat, mesh.newQuat, angleBetween * f)*

There are a few traits of mesh that refer to the *before* and *after* key frame states, in addition to the mesh's current state (between key frames). *mesh.oldPosition* represents the mesh's coordinates in the current key frame, and *mesh.newPosition* represent the mesh's coordinates in the next key frame. *mesh.position* of course refers to the mesh's current position, whether it is on a key frame or in between key frames. The same convention also goes for the mesh's rotational states, expressed in quaternions. There are several Unity provided functions that were used in the algorithm. Instead of having to create a matrix and quaternion library with all of its functionalities from scratch, there were already useful functions provided, such as *rotateTowards*. That function would return a Quaternion data type which was determined by rotating a point in space around a vector by a specified angle.

6.2.2 Increment

Referring to the above pseudo-code, each mesh in the arm gets incremented, via translation and rotation, towards its destination state. The size of this increment determined the speed of the animation itself, assuming a constant 60 frame per second game clock. The size of the increment is determined by a proportional value f, which is multiplied by the total magnitude between each state to get the size of the increment value. The parameter f was tweaked to get a smooth but quick enough animation, until deemed satisfactory.

CHAPTER 7
CONCLUSION

This thesis sought to explore the objective and subjective differences between ego-sensitive and non-ego-sensitive allocentric mixed reality deictic gestures. As hypothesized, we discovered a dichotomy between these two gestural categories that presents a challenge for robot designers. Specifically, while ego-sensitive allocentric gestures such as pointing with virtual arms result in social benefits such as increased social presence, perceived anthropomorphism, and likability, non-ego-sensitive allocentric gestures such as virtual arrows result in greater task performance with respect to both speed and accuracy. In future work we plan to explore whether robots may achieve the "best of both worlds" by using both visualizations together, or whether this would be too cognitively overloading or perceived as too busy.

While our secondary distance-oriented hypotheses were not supported, for several of our objective and subjective measures our analyses were unable to directly support *or* refute these hypotheses, suggesting that more data must be collected before a decision can be made one way or another. Because our experiment was conducted using a Bayesian analysis framework, we are able to do just this, without violating a sampling plan or test assumptions, and thus plan to do so in future work.

According to the data in our experiment, we have made significant progress in better understanding the various effects of transitioning traditional robotic deictic gestures into a virtual world, and exploring what different types have to offer when they are manifested in a mixed reality environment. However, there still remains a dichotomy of methods we used in this experiment. Our non-ego-sensitive allocentric gesture provides promise in situations where a bond between a human and a robot is not needed. However, in contexts where a bond is a necessary component for trust and robot dependence, there needs to be a viable solution for obtaining maximum performance and positive social perception.

Another possible limitation of resorting to using a non-ego-sensitive gesture, is in situations where the human-robot team consists of more than one gesturally capable robots. As mentioned in Section 1, non-egocentric allocentric gestures, such as drawing referent arrows, redirects focus from the robot to the disconnected gesture itself. When this phenomena occurs in the context of multiple robots, the human team members will have no clear way of tracing the gesture back to its source, due to the disconnect that these gestures incur. There could be some way of fighting against these consequences, such as color-coding the gestural display on the users head mounted display, or perhaps connecting a virtual tether between the robot and the gesture it allocates. In future experiments, it could be possible to examine these limitations of the same mixed reality deictic gestures in contexts with more than one gesturing robots. However, due to the computational limitations of most head mounted displays, multiple sources of mixed reality gestures that are controlled with computationally heavy computer vision algorithms may prove to be too taxing on said devices.

CHAPTER 8

ACKNOWLEDGMENTS

This work was funded in part by National Science Foundation grants IIS-1909864 and CNS-1810631.

REFERENCES CITED

[1] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[2] Adrian Bangerter and Max M Louwerse. 2005. Focusing attention with deictic gestures and linguistic expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 27.

[3] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.

[4] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 708–713.

[5] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 297–304.

[6] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 297–304.

[7] Judee K Burgoon and Gregory D Hoobler. 1994. Nonverbal signals. *Handbook of interpersonal communication* 2 (1994), 229–285.

[8] Rehj Cantrell, Paul Schermerhorn, and Matthias Scheutz. 2011. Learning actions from human-robot dialogues. In *2011 RO-MAN*. IEEE, 125–130.

[9] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS) Development and Validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. 254–262.

[10] Elizabeth Cha, Yunkyung Kim, Terrence Fong, Maja J Mataric, et al. 2018. A survey of nonverbal signaling methods for non-humanoid robots. *Foundations and Trends® in Robotics* 6, 4 (2018), 211–323.

[11] Noel K Clark and Derek R Rutter. 1985. Social categorization, visual cues, and social judgements. *European Journal of Social Psychology* 15, 1 (1985), 105–119.

[12] Susan Wagner Cook, Terina Kuangyi Yip, and Susan Goldin-Meadow. 2012. Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and cognitive processes* 27, 4 (2012), 594–610.

[13] Martin J Crowder. 2017. *Analysis of repeated measures*. Routledge.

[14] Antonella De Angeli, Walter Gerbino, Giulia Cassano, and Daniela Petrelli. 1998. Visual display, pointing, and natural language: the power of multimodal interaction. In *Proceedings of the working conference on Advanced visual interfaces*. 164–173.

[15] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological review* 114, 4 (2007), 864.

[16] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.

[17] Susan Goldin-Meadow. 1999. The role of gesture in communication and thinking. *Trends in cognitive sciences* 3, 11 (1999), 419–429.

[18] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J Matarić. 2019. Using Socially Expressive Mixed Reality Arms for Enhancing Low-Expressivity Robots. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–8.

[19] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics* 2, 4 (2010), 361–375.

[20] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of computer-mediated communication* 3, 2 (1997), JCMC321.

[21] William Marslen-Wilson, Elena Levy, and Lorraine K Tyler. 1982. Producing interpretable discourse: The establishment and maintenance of reference. *Speech, place, and action* (1982), 339–378.

[22] S. Mathôt. 2017. Bayes like a Baws: Interpreting Bayesian repeated measures in JASP [Blog Post]. https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp.

[23] Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.

[24] Christoforos I Mavrogiannis, Minas V Liarokapis, and Kostas J Kyriakopoulos. 2015. Quantifying anthropomorphism of robot arms. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4084–4089.

[25] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. 1993. Applications of augmented reality for human-robot communication. In *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*, Vol. 3. IEEE, 1467–1472.

[26] RD Morey and JN Rouder. 2014. BayesFactor (Version 0.9. 9).

[27] Raedy Ping and Susan Goldin-Meadow. 2010. Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science* 34, 4 (2010), 602–619.

[28] Joanne Pransky. 2001. AIBO–the No. 1 selling service robot. *Industrial robot: An international journal* (2001).

[29] Patrick Renner, Florian Lier, Felix Friese, Thies Pfeiffer, and Sven Wachsmuth. 2018. Facilitating HRI by mixed reality techniques. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 215–216.

[30] Laurel D Riek, Philip C Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces* 3, 1-2 (2010), 99–108.

[31] Tina L Robbins and Angelo S DeNisi. 1994. A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations. *Journal of Applied Psychology* 79, 3 (1994), 341.

[32] Jeffrey N Rouder, Richard D Morey, Paul L Speckman, and Jordan M Province. 2012. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology* 56, 5 (2012), 356–374.

[33] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5, 3 (2013), 313–323.

[34] Maha Salem, Stefan Kopp, Ipke Wachsmuth, Katharina Rohlfing, and Frank Joublin. 2012. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics* 4, 2 (2012), 201–217.

[35] Allison Sauppé and Bilge Mutlu. 2014. Robot deictics: How gesture and context shape referential communication. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 342–349.

[36] Matthias Scheutz, Paul Schermerhorn, James Kramer, and David Anderson. 2007. First steps toward natural human-like HRI. *Autonomous Robots* 22, 4 (2007), 411–423.

[37] Takanori Shibata, Kazuyoshi Wada, and Kazuo Tanie. 2003. Statistical analysis and comparison of questionnaire results of subjective evaluations of seal robot in Japan and UK. In *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, Vol. 3. IEEE, 3152–3157.

[38] Adam Stogsdill, Thao Phung, and Tom Williams. 2020. Investigating Confidence-Based Category Transition of Spatial Gestures. In *Proceedings of the 2nd Workshop on Natural Language Generation for Human-Robot Interaction at HRI 2020.*

[39] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2015. Communicating directionality in flying robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 19–26.

[40] JASP Team. 2018. JASP (Version 0.8.5.1)[Computer software].

[41] Nhan Tran, Kai Mizuno, Trevor Grant, Thao Phung, Leanne Hirshfield, and Tom Williams. 2019. Exploring Mixed Reality Robot Communication Under Different types of Mental Workload. In *Proceedings of the 3rd International Workshop on Virtual, Augmented, and Mixed Reality for HRI.*

[42] Albert van Breemen, Xue Yan, and Bernt Meerbeek. 2005. iCat: an animated user-interface robot with personality. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. 143–144.

[43] EJ Wagenmakers, J Love, M Marsman, T Jamil, A Ly, and J Verhagen. 2018. Bayesian inference for psychology, Part II: Example applications with JASP. *Psychonomic Bulletin and Review* 25, 1 (2018), 35–57.

[44] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed Reality Deictic Gesture for Multi-Modal Robot Communication. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction.*

[45] Tom Williams, Matthew Bussing, Sebastian Cabrol, Ian Lau, Elizabeth Boyle, and Nhan Tran. 2019. Investigating the Potential Effectiveness of Allocentric Mixed Reality Deictic Gesture. In *Proceedings of the 11th International Conference on Virtual, Augmented, and Mixed Reality.*

[46] Tom Williams and Matthias Scheutz. 2017. Referring Expression Generation Under Uncertainty: Algorithm and Evaluation Framework. In *Proceedings of the 10th International Conference on Natural Language Generation.*

[47] Tom Williams, Daniel Szafir, and Tathagata Chakraborti. 2019. The Reality-Virtuality Interaction Cube. In *Proceedings of the 2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI.*

[48] Tom Williams, Nhan Tran, Josh Rands, and Neil T. Dantam. 2018. Augmented, Mixed, and Virtual Reality Enabling of Robot Deixis. In *Proceedings of the 10th International Conference on Virtual, Augmented, and Mixed Reality.*