Communicating Robot Motion Intent with Augmented Reality

Michael Walker	Hooman Hedayati	Jennifer Lee	Daniel Szafir
University of Colorado Boulder	University of Colorado Boulder	University of Colorado Boulder	University of Colorado Boulder
Colorado, United States	Colorado, United States	Colorado, United States	Colorado, United States
michael.walker-1@colorado.edu	hooman.hedayati@colorado.edu	jennifer.c.lee@colorado.edu	daniel.szafir@colorado.edu

Figure 1: In this work, we explore how augmented reality might mediate collocated human-robot interactions by visually conveying robot motion intent. We developed four reference prototypes for cuing aerial robot flight motion: (A) NavPoints, (B) Arrows, (C) Gaze, (D) Utilities.

ABSTRACT

Humans coordinate teamwork by conveying intent through social cues, such as gestures and gaze behaviors. However, these methods may not be possible for appearance-constrained robots that lack anthropomorphic or zoomorphic features, such as aerial robots. We explore a new design space for communicating robot motion intent by investigating how augmented reality (AR) might mediate human-robot interactions. We develop a series of explicit and implicit designs for visually signaling robot motion intent using AR, which we evaluate in a user study. We found that several of our AR designs significantly improved objective task efficiency over a baseline in which users only received physically-embodied orientation cues. In addition, our designs offer several trade-offs in terms of intent clarity and user perceptions of the robot as a teammate.

CCS CONCEPTS

• **Computer systems organization** → **External interfaces for robotics**; • **Human-centered computing** → **Mixed / augmented reality**; **User interface design**; **Laboratory experiments**;

KEYWORDS

Aerial robots; robots; drones; robot intent; augmented reality; mixed reality; ARHMD; interface design; virtuality continuum;

1 INTRODUCTION

Effective collaboration requires that teammates quickly and accurately communicate their intentions to build common ground, coordinate joint actions, and plan future activities. For example, prior work in social, cognitive, and behavioral sciences has found that collaborative activities fundamentally depend on *interpredictability*—the ability of each team member to rapidly understand and predict their teammate's attitudes and actions [23]. In collocated human-robot teams, poor communication of robot intentions and planned movements can lead to critical breakdowns that degrade safety, task performance, and perceptions of robot usability.

As a result, providing support for this *motion inference problem*, where users may have difficulties understanding when, where, and how a robot teammate will move, represents a primary challenge towards achieving safe and usable robotic systems. In human-human teams, people use a variety of implicit and explicit cues, such as gaze, gestures, or other social behaviors, that communicate planned actions and movements to enhance team effectiveness and help maintain trust. Research has demonstrated that robots may also use social cues to convey both movement intent and affective state [12, 27, 36]. However, it is not always clear how to apply these findings to robots that lack anthropomorphic and zoomorphic features, such as industrial robot manipulators or aerial robotic free-flyers.

Instead, recent work in HRI has begun to explore alternative methods for supporting motion inference, including generating legible motion trajectories [10], developing expressive motion primitives [42], verbalizing robot intentions using natural language [46], using projector-based or electronic display systems to provide additional information [2, 7, 22, 30, 37, 38, 48], and using light signals as explicit directionality cues [5, 43]. While such advances have shown promise in enhancing interaction safety and fluidity, a variety of constraints arising from environmental, task, power, computational, and platform considerations may limit their feasibility or effectiveness in certain contexts. For example, altering robot motions for legibility or expressiveness may not always be possible in dynamic or cluttered environments, natural language may not be a practical form of feedback in noisy environments (e.g., manufacturing warehouses or construction sites) or for robotic platforms that generate

a great deal of noise (e.g., aerial robots), and projections may be difficult to render on non-flat surfaces, may not be salient in bright environments, and can be occluded by user or robot.

In this paper, we explore an alternative design space: using augmented reality to resolve motion inference. Our work is inspired by prior research envisioning that augmented reality head-mounted displays (ARHMDs) might one-day support intuitive human-robot communication for collocated users [8, 16, 17, 26, 35]. While past efforts integrating robotics and augmented reality were hampered by limitations in underlying AR technologies (often relying on custom-built HMDs with limited fidelity), recent industry developments are increasing the availability of standardized, consumer-grade ARHMD technology capable of providing rich, intuitive, visual feedback. These advances are providing an opportunity for an exciting new design space: HRI mediated by hands-free, see-through augmented reality. We provide the first in-depth examination of this design space within the context of communicating robot motion intent to collocated users, specifically for aerial robots.

2 BACKGROUND

In this work, we build on prior research focused on resolving robot motion inference by exploring a new design space where augmented reality cues communicate robot intent. Below, we review related work on communicating robot intent and provide a brief primer on augmented reality technologies.

2.1 Communicating Robot Intent

The robot motion inference problem may be considered as an analog to the "gulf of evaluation" [29] issue that commonly arises between representations provided by a system and user abilities to interpret the system [43]. This can be especially challenging for robots with high degrees of freedom, such as aerial robots. Other issues may compound this problem, including a lack of robot capabilities for communicating intention and goals using traditional methods and technological novelty/lack of mature mental models for understanding robot behaviors.

Previous research has suggested that effective robot communication enhances user perceptions of the robot's reliability, predictability, and transparency and may increase user willingness to accept and use new robotic technologies in work environments [7]. Research has also shown that cuing robot intent can help users anticipate and predict robot directional motion faster, enabling them to respond more quickly in interactive tasks [14] while increasing user preferences for working with robots [42]. Legible motion that expresses a robot's intentions can further improve interaction fluidity and efficiency in collaboration between humans and robots [10].

Past research in robot design has examined how to effectively leverage users' prior experiences and mental models in human-human collaboration to bootstrap human-robot collaboration, imbuing robots with social behaviors such as gaze and gestures that people commonly use. Such behaviors have been explored for a variety of robots using anthropomorphic and zoomorphic features (e.g., [3, 14, 18, 27, 45]). For aerial robots that lack such features, prior work has examined expressive flight patterns, demonstrating that certain behaviors based on biological motion and principles from film and animation may help offset the lack of developed mental

models for free-flying movement [42]. Research has also explored more explicit cues, such as using lights as indicators [5, 43] and mixed-reality projection systems [2, 7, 48], finding that the use of projected imagery can be advantageous for communicating spatial intentions and instructions such as informing human collaborators of robots' intended path [48].

Although such work has shown promising benefits for improving human-robot interaction, including with aerial robots, prior methods are not without limitations. For example, expressive flight motions may not be feasible in constrained environments or effective if the user simply glances at the robot and witnesses only part of the motion. Projection systems rely on instrumenting the environment and face difficulties at distance, in chaotic environments, and run the risk of occlusion. In this work, we explore AR as a means of conveying robot motion intent in a manner not bound by these limitations.

2.2 Augmented Reality

Augmented reality (AR) is a technology that overlays computer graphics onto real world environments in real-time [16, 17]. AR interfaces have three main features: (1) users are able to view both real and virtual objects in a combined scene, (2) users receive the impression that virtual objects are actually expressed and embedded directly in the real world, and (3) the virtual objects can be interacted with in real-time [4]. This contrasts with purely virtual environments or other parts of the virtuality continuum (VC), such as augmented virtuality in which real-world objects are mixed into a virtual environment [25].

AR technologies show promise in supporting human activities across a variety of domains. While early systems, starting with Sutherland's "Sword of Damocles" prototype [39], were quite limited in display fidelity, rendering speed, support for interaction, and generalizability, recent advancements in augmented reality head-mounted display (ARHMD) technology is creating an ecosystem of standardized, consumer-grade see-through ARHMDs. For example, the HoloLens and Meta 2 ARHMDs both afford high resolution stereographic virtual imagery displayed at a 60Hz refresh rate, built-in gesture tracking, depth sensing, and integration with standard development tools such as Unity and Unreal Engine. This advance in hardware accessibility is creating new opportunities for exploring AR as an interaction medium for enhancing HRI.

Although we are not the first to recognize that AR holds potential for improving human-robot interactions (see [8, 16, 17, 26, 35]), we believe that this area is critically understudied and represents a fairly nascent research space, especially within the context of the capabilities provided by modern ARHMD hardware. We think that AR may be more flexible than past solutions that rely on physically-embodied cues, flight patterns, or projections and might be easily adapted to a variety of tasks and environments. For example, AR might overcome limitations arising from complex environments by superimposing visuals in a head-mounted display without limiting robot mobility or risking occlusion [26].

3 DESIGN PROCESS

We undertook an iterative design process to explore the space of how ARHMD technology might mediate collocated human-robot

interaction by providing feedback on robot motion intent. Our design process began with an analysis of the potential of augmented and mixed reality technology. Synthesizing information from past work across the VC, including mixed-reality projection systems, AR entertainment applications, and augmented virtuality educational software, we developed a high-level framework for considering how ARHMDs might enhance human-robot interactions.

3.1 A Design Framework for AR-HRI

Our framework classifies potential designs for augmenting human-robot interactions with virtual imagery into three main categories, regarding whether additional information is communicated to the user by (1) *augmenting the environment*, (2) *augmenting the robot*, or (3) *augmenting the user interface*. This framework of broad design concepts allows us to survey the landscape of possible AR interfaces and provides a structure for reasoning about requirements, benefits, and trade-offs in designing AR for HRI.

Augmenting the environment: In this paradigm, virtual imagery is represented as new cues directly embedded into the context of the shared work area using an *environment-as-canvas* metaphor. This notion extends past work using mixed-reality projection systems in three major ways. First, ARHMDs afford a full three-dimensional "canvas" to utilize for virtual imagery, rather than a two-dimensional canvas from projector systems. Second, ARHMD environmental augmentations do not risk being occluded, as when a user or robot interferes with projected light. Third, ARHMDs support stereographic environmental cues that can better leverage human depth perception, as opposed to monocular cues in traditional projector systems.

Augmenting the robot: In this archetype, virtual imagery is directly connected to the robot platform to alter robot morphology in a *robot-as-canvas* metaphor. This technique may alter robot form and/or function by creating new "virtually/physically embodied" cues, where cues that are traditionally generated using physical aspects of the robot are instead generated using indistinguishable virtual imagery. For example, rather than directly modifying a robot platform to include signaling lights, as in [43], an ARHMD interface might overlay virtual signaling lights on the robot in an identical manner. Alternatively, virtual imagery might be used to give anthropomorphic or zoomorphic features to robots that don't have this physical capacity (e.g., adding a virtual body to a single manipulator or a virtual head to an aerial robot). Finally, virtual imagery might be used to obscure or make more salient various aspects of robot morphology based on user role (e.g., an override switch might be hidden for normal users but visible for a technician). Overall, we believe this novel paradigm, in which robot morphology becomes a design variable that is fast, easy, and cheap to prototype and manipulate, opens up a great deal of exciting new potential for HRI researchers and designers, especially since modifications to robot morphology are traditionally prohibitive due to cost, time, and/or task/environment constraints.

Augmenting the user interface: In this paradigm, virtual imagery is provided directly in front of the user, giving them an interface to the physical world, inspired by "window-on-the-world" AR applications [11] and heads-up display technologies used for pilots [13, 24]. This *interface-as-canvas* metaphor may uniquely supply egocentric cues, either directly in front of the user's view or in their periphery,

Table 1: Visual summary of features supported by each design.

	NavPoints	Arrow	Gaze	Utilities
Display Category	Augmented Environment	Augmented Environment	Augmented Robot	Augmented User Interface
Next Destination Direction	Explicit	Explicit	Explicit	
Next Destination Location	Explicit	Explicit		
Future Destination Location	Explicit	Explicit		
Arrival Time	Explicit	Inferred		
Wait Time	Explicit	Inferred	Explicit	
Velocity	Explicit	Inferred		
Relative Position				Explicit

compared to the exocentric feedback provided by augmenting the environment or robot [47]. For example, user interface augmentations might include spatial mini-maps that provide information on the position or planned route of robots relative to the user, robot status indicators (battery level, task progress, task queue, etc.), or live video streams from a robot's camera(s).

3.2 Creating Reference Prototypes

Inspired by prior work in communicating robot intent, we used the design framework described above to develop several reference designs that sample from the potentially infinite solution space of AR visualizations for conveying robot intent. Our initial designs ranged from showing dynamic virtual fences that outlined "safe" and "unsafe" areas in the environment, to adding virtual engines to the robot (inspired by [43]), to various ways of encoding robot motion trajectories, velocities, and accelerations. After conducting series of design iterations based on feedback from short pilot tests, in which users viewed a collocated aerial robot while using a preliminary ARHMD design prototype, we eventually narrowed our candidate prototypes to four main designs.

We refer to these four candidate designs as *NavPoints*, *Arrows*, *Gaze*, and *Utilities* (Figure 1). Together, they sample from each paradigm in our design framework and offer potential trade-offs in terms of information conveyed, information precision, generalizability, and possibility for distraction/interface overdraw. Table 1 provides a visual summary of features supported by each design.

While our work examines these designs from the context of conveying robot motion intent for aerial robots, we believe that our design framework and methodology, along with the main design metaphors and interface techniques we develop, may generalize to additional robotic platforms where conveying intent may facilitate human-robot interaction (e.g., industrial robotic manipulators that also move with high degrees-of-freedom). Note the following designs have heuristic parameters (e.g., X, Y, Z) able to be tuned to the specifications of the designer. We provide the values we chose for our specific implementations of these designs in §4.2.

NavPoints: The NavPoints design is an example of augmenting the environment. This design has a spatial focus as it provides virtual imagery that displays the robot's planned flight path as a series of X lines and navigation waypoints, similar to what might be seen in traditional waypoint-delegation or supervisory interfaces (e.g., [44]). The lines connect the robot's current position to its future destinations in sequential order. Destination waypoints are visualized as spheres and indicate the robot's precise destinations in 3D space. Each destination sphere also renders a drop shadow on the ground directly below it, which has been shown to aid in depth estimations made by the user [9]. Above each navigation point are two radial timers. An inner white timer indicates when the drone will arrive at that location and the outer dark blue timer indicates when the robot will leave that location. Smaller spheres travel along the lines, moving in the same direction and velocity the robot will travel between destinations, thus providing an anticipatory cue for future robot velocity and direction. Information displayed by this design regarding velocity and arrival/departure timings is thus explicitly displayed for the user.

Arrow: The Arrow design provides an alternate example of how virtual imagery might augment a shared environment. While the NavPoints design provides users with a great deal of information, it may be distracting or confusing to the user due to potential overdraw. The Arrow design takes a more minimal approach, focusing specifically on communicating temporal information inspired by [8] and common user experiences with modern GPS systems. The virtual imagery consists of a blue arrow head that travels through 3D space in the exact path the robot will eventually take X seconds in the future. As the arrow moves, a line is left behind that traces the arrow's path back to the robot. Using this line, users can explicitly see the path the arrow has taken, which the robot will eventually follow. The line the arrow creates renders a drop shadow on the ground directly beneath it. Information displayed by this design regarding velocity and arrival/departure timings must be inferred by the user by watching the discrete movements of the arrow through space.

Gaze: The Gaze design represents an example of augmenting the robot itself with virtual imagery. This design is inspired by prior research that has demonstrated the remarkable potential of gaze behaviors for conveying intent, even for aerial robots [43], as well as prior designs for robotic blimps [19] and research in robotic telepresence that has explored a metaphor of treating an aerial robot as a "floating head" [20]. This design provides virtual imagery that completely alters the robot form by overlaying a X-meter diameter white sphere with a pupil directly over the aerial robot, effectively transforming the robot from a multirotor into a "flying eye." While moving between destinations, the eye model stares at its current destination until it enters a predetermined distance threshold of Y meters between itself and the current destination, at which point the eye turns and focuses on the robot's next destination. Gaze shifts such as this have been shown useful in predicting action intent in humans [21]. Through these shifts in focus, the robot's immediate destinations are preemptively revealed to the user. If the robot is to remain stationary at a destination for longer than Z seconds, the lens over the normally transparent pupil becomes opaque. When the now stationary robot is within Z seconds of departing, the lens fades in back to transparent. This fade in is done as a linear interpolation over the course of Z seconds. This lens fade in/out effect notifies the user how long the robot will remain stationary and was inspired by *accommodation*—shifts as ciliary muscles contract and relax in human gaze as focus switches between near and far targets—and lens focusing in traditional cameras. The size of the display was chosen to help users more easily determine gaze direction at near and far distances from the robot. The back of the sphere directly behind the pupil is rendered as flat to help users infer the rotation of the eye when not facing the robot head on. Finally, the eye casts a drop shadow on the ground directly below it.

Utilities: Our final design illustrates a potential method for augmenting a user interface to provide contextual information in an egocentric manner. This design is inspired by peripheral utilities, such as minimaps, radar, and off-screen indicators that often augment pilot interfaces [24], robot control interfaces [28], videogame interfaces [32, 34], and military applications [15]. This design displays a 2D circular "radar" fixed at the bottom left corner of the ARHMD display. The user is displayed as a blue dot that is always centered within the radar, while the robot is rendered as a red dot on the radar relative to the user's location. The size of the robot's radar dot is directly proportional to its current height. The detection radius, X, of the radar can be customized by the interface designer or adjusted by the user. When the robot is in the user's field-of-view (FOV), it is overlaid with a targeting box, when not in the FOV, an off-screen indicator appears in the form of a directional arrow. This arrow is rendered along the side of the ARHMD display pointing to the location of the off-screen robot. Both radar and targeting box/off-screen indicator provide the user with the means to rapidly locate the robot relative to themselves.

4 EXPERIMENT

We conducted a 5×1 between-participants experiment to evaluate how our designs affect user interactions with a flying robot in a shared workspace. The independent variable in this study was the type of AR feedback the user received (five levels: a baseline and our four designs described above). In the baseline condition, participants still wore an ARHMD, but did not see any virtual imagery. Instead, participants in this condition were informed that the robot has a distinct "front," which always indicates its direction of flight; this baseline behavior meant the robot would always orient itself to the direction of travel, leveraging the only physically-embodied cue that the robot's default morphology provides (Figure 3). All conditions shared this baseline orientation behavior. Dependent variables included objective measures of task performance and efficiency as well as subjective ratings of communication clarity and robot usability.

4.1 Task and Environment

Our overall experimental setup was inspired by contexts where free-flying robots might assist with logistics management in manufacturing settings in the near future [40]. In the study, participants worked with an aerial robot in a shared environment designed to mimic a small warehouse. The environment measured $20ft \times 35ft \times 20ft$ and contained motion tracking cameras that we utilized for precise robot navigation to ensure participant safety.

Six workstations were placed within the physical space in two rows of three (Figure 2). Each workstation had at least five feet of surrounding free space and supported a container of colored beads.
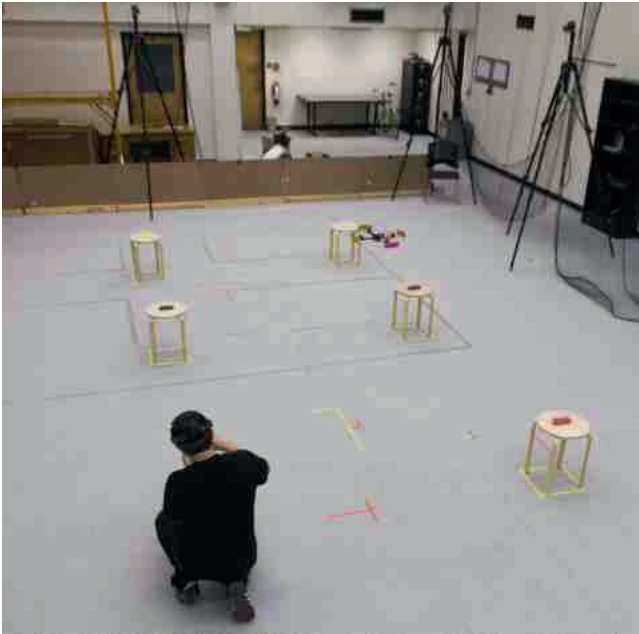
Figure 2: The experimental environment required that participants share a workspace with a collocated arial robot.

Each bead container held only one color of beads, corresponding to either green, black, yellow, white, blue, or red. Participants were tasked with collecting beads from these containers and fastening them together to make beaded strings while sharing the environment with an aerial robot. Participants were instructed that their goal was to make as many beaded strings as possible in exactly eight minutes. Each completed string consisted of twenty-five beads. There were individual instructions for each string describing the target color and amount of beads to be used. For example, one string might ask for 10 blue beads, 5 red beads, and 10 green beads. Along with these directions, participants were instructed on three additional rules:

1. Participants could only pick up one bead at a time and could only place beads on strings while at the workstation.
2. Participants could collect the colors in any order, but once they chose a color they had to remain at that color station until they had strung all the beads of that color, as indicated by the string instructions (i.e., colors could not be intermixed).
3. The robot would occasionally visit each workstation (ostensibly to monitor bead supply). If the robot flew to the workstation where participants were working, participants were required to move at least 2 meters away from the workstation (i.e., moving back to *social distance* as informed by proxemics [1]) and wait until the robot left before continuing (i.e., giving workstation priority to the robot).

This task was designed to emulate an assembly task that might be found in a warehouse, with shared resources (i.e., the workstations) between the user and robot. As the robot had priority at workstations, the task required that participants understand and predict robot intent to best plan their activities and maximize task efficiency.

4.2 Experimental Apparatus & Implementation

Robotic Platform: We used an AscTec Hummingbird robot as our free-flying platform (Figure 3). During the experiment, the robot flew autonomously to pre-programmed waypoints throughout the experimental environment using a PID controller that received input on robot location using a motion capture system. During the study, a researcher stood by with an emergency kill switch that could disarm the robot for safety, but this was never required.

ARHMD Platform: We used a Microsoft HoloLens as our ARHMD. The HoloLens is a wireless, optical see-through stereographic augmented reality HMD with a $30° \times 17.5°$ FOV, an inertial measurement unit, depth sensor, ambient light sensor, and several cameras and microphones that support voice input, gesture recognition, and head tracking. The HoloLens was chosen due to its emerging popularity, ease of access, ability to support hands-free AR, and high potential as a model for future consumer ARHMD systems.

Experimental Framework: We developed a custom experimental framework for implementing our designs, deploying them to the HoloLens, and ensuring that visualizations are properly synchronized with robot behaviors. We first prototyped our four designs described in §3.2 using Unity, a popular game and animation engine for designing and developing virtual and AR applications. We also developed a waypoint system that enables the specification of a sequential list of desired robot destinations (i.e., target robot positions and orientations in 6 degree-of-freedom space), desired velocity to travel to each destination, and wait time at each destination (possibly zero). We then added an invisible *virtual drone* object to the Unity scene that navigates the scene per the specifications of the waypoint list, and whose motions control the physical robot's flight. For every iteration of the game engine's update loop, we send the virtual drone's 3D pose from the HoloLens to the backend system controlling the physical robot via UDP. Virtual drone pose values are transformed from Unity coordinates to real-world coordinates using a transformation matrix calculated *a priori* in an initial calibration procedure. After this transformation, pose values in the Unity scene correspond to the same positions in the user's environment, enabling the physical robot to fly through the study space in a manner that matches the virtual drone, turning our Unity application into a waypoint delegation interface as in [44]. This framework allows us to calculate the exact position of the physical robot relative to the user at any time and adjust our displayed visualizations accordingly.

For the experiment, the waypoint list specified in Unity consists of destinations corresponding to the six workstations. The waypoints were placed directly above the workstations between 1–1.5 meters high. Upon arriving at a workstation, the robot would pause for 5–15 seconds to simulate monitoring bead inventory. The robot's velocity when traveling between stations ranged from 0.2–1m/s, as suggested by prior research into appropriate robot speeds for collocated interaction [6, 31]. The waypoint list was populated by randomly selecting workstation destinations, wait times, and velocities from the ranges listed above to ensure there was no flight path pattern for participants to learn. This list was populated once prior to the start of the study as a whole, then held constant for all participants across all conditions to remove potential variance arising from different robot trajectories. The total time to travel through the waypoint list took 8 minutes to

Figure 3: Participant making a bead string at one of the six assembly stations mid-task. The AscTec Hummingbird robot flies nearby (colored corner marks robot "front" for baseline participants).

complete, which was equivalent to the total task duration (chosen due to battery/flight time capabilities).

We also tuned the heuristic parameters in implementing the designs in §3.2. NavPoints has the potential to display a full flight path of all waypoints, but early pilot tests revealed this led to a distracting case of overdraw. We instead opted to show only a portion of the robot's path, which dynamically updated such that the visualization always displayed three waypoint destinations ahead of the robot's current position. The Arrow design was implemented such that the blue arrow always stayed fifteen seconds ahead of the aerial robot (i.e., showing the robot's position $15s$ in the future). The heuristic parameters selected for eye movement in the Gaze design were derived from research into human gaze behaviors [21] and set to a distance threshold of $2m$ and fade threshold of $3s$. The radius of the eye model was set to $0.6m$ so that the hologram overlaid the robot's body in its entirety. The "radar" in the Utilities design used a detection radius of $15m$ to cover the full area of the experiment environment. Final implementations of these designs can be seen in Figure 1 and in our supplemental video.

4.3 Participants & Procedure

We recruited a total of 60 participants (40 males, 20 females, evenly balanced across conditions) from the University of Colorado Boulder campus to take part in our study, which was approved by our university IRB. Average participant age was 20.7 (SD = 4.8), with a range of 18–45. On a seven-point scale, participants reported a moderate prior familiarity with aerial robots (M = 3.75, SD = 1.71), but a low familiarity with ARHMDs (M = 2.65, SD = 1.83).

The study took approximately 30 minutes and consisted of five phases: (1) introduction, (2) training, (3) calibration, (4) task, and (5) conclusion. (1) First, participants signed a consent form and were led into the task space. (2) Next, participants each read identical instruction sheets detailing the task and task rules. Participants assigned to any of the four AR design conditions described in §3.2 then watched a corresponding $60s$ tutorial video that provided a brief instruction on the AR feedback they would receive based on the relative novelty of ARHMD technology. Only participants assigned to the baseline condition were told verbally that the robot always

moves towards where its marked "front" is facing, as described in §4. (3) The ARHMD application was then started, calibrated, and fitted on each participant, including participants assigned to the baseline condition (even though they didn't receive AR feedback). (4) Participants then performed the main task for eight minutes, making as many bead strings as possible while sharing the environment with a collocated aerial robot, as described in §4.1. (5) Once eight minutes of task time was completed, participants were told to stop and were given a post-survey on their experience.

4.4 Measures & Analysis

We used a combination of four objective and subjective measurements to characterize the effects of our designs. Objective *task efficiency* was measured by the total time participants spent waiting while interrupted by the robot, which could be avoided by planning their work with an understanding of robot intent (lower times indicate better performance/efficiency). In calculating task efficiency, we removed time variance in users backing away from the stations. The interruption timer only began when the robot was directly overhead a station. This allowed for consistent measuring between participants.

We also constructed a number of scales from 7-point Likert-style questionnaire items to measure subjective participant perceptions and preferences. Scales rated interface design *clarity* (4 items, Cronbach's α = .85), perceptions of the robot as a *teammate*, both as a personal work partner (4 items, Cronbach's α = .91), and as a potential work partner for others (2 items, Cronbach's α = .83), and overall *design usability* (2 items, Cronbach's α = .73).

Qualitative feedback was obtained through open-ended questions posed to each participant as part of the concluding questionnaire to describe their experiences "working with the AR user interface", "working alongside the aerial robot", and "completing [their] task".

We analyzed data using a one-way Analysis of Variance (ANOVA) with experimental condition (i.e., interface design) as a fixed effect. Post-hoc tests used Dunnett's method to control for Type I errors when evaluating the ARHMD designs against the baseline condition, while Tukey's Honestly Significant Difference (HSD) test compared effectiveness across each design.

5 RESULTS

Figure 4 summarizes our main objective and subjective results.

Objective Results — We analyzed our task performance metric to confirm that our designs were useful for participants to quickly and accurately deduce robot intent and plan their own activities more effectively. We found a significant main effect of ARHMD interface design on total time spent interrupted, $F(4, 55) = 12.56$, $p < .001$. Comparing the performance of each design to the baseline with Dunnett's multiple comparison test, we found that total time lost to interruptions significantly decreased using NavPoints ($p < .001$), Arrow ($p < .001$), and Gaze ($p = .003$), but not Utilities ($p = .104$).

Subjective Results — Participants rated the several facets regarding the communication of robot movement intent. We found a significant effect of design on perceived communication clarity, $F(4, 55) = 11.04$, $p < .001$. Post-hoc comparisons using Dunnett's test revealed that the NavPoints design was rated significantly higher than the baseline ($p < .001$), but we did not find significant effects from the other designs.

We analyzed participant responses to the robot in terms of how they might view it as a collaborative partner in a work environment. We found a marginal $(0.1 > p > 0.05)$ main effect of design on participant perceptions of the robot as a good work partner for themselves, $F(4, 55) = 2.48$, $p = .054$. We also found a significant main effect of design on participant perceptions of the robot as a good work partner for others, $F(4, 55) = 2.54$, $p = .049$. Post-hoc comparisons revealed NavPoints was the only design to significantly improve perceptions of the robot as a personal work partner ($p = .03$) and as a work partner for others ($p = .029$) over the baseline.

Finally, we compared the designs to one another along a usability metric of how the displayed virtual imagery affected participant understanding of robot movement intent. We found a significant main effect of design on perceived usability for understanding intent, $F(3,44) = 25.32$, $p < .001$. Post-hoc comparisons using Tukey's HSD found that NavPoints ($M = 6.96$), $p < .001$, Arrow ($M = 6.67$), $p < .001$, and Gaze ($M = 5.83$), $p < .001$, were ranked as significantly more helpful than Utilities ($M = 4.21$). We also found that NavPoints was rated as significantly more helpful than Gaze, $p = .012$, with Arrow ranked marginally more helpful than Gaze, $p = .092$.

6 DISCUSSION

The NavPoints, Arrow, and Gaze designs improved task performance by reducing inefficiency; participants were able to better predict robot intent and plan their own actions accordingly to reduce the length of times they spent interrupted and unproductive. However, the Utilities model did not provide similar improvements over the baseline condition. This may be due to the Utilities design emphasizing current robot positioning relative to the user rather than displaying cues that help users predict the robot's future destinations, unlike the other designs. Participant responses support this conclusion, revealing similarities between baseline and Utilities participants:

P16 [Utilities]: *"I wasn't able to predict when the drone would move or if there was a pattern. But when it did start moving, I could figure out where it was going and could plan my actions accordingly."*
P02 [Baseline]: *"It was difficult to tell where the drone was going go and for how long it would stay there. . . often times you had no way of knowing what it was going to do."*

Another possible reason for the poorer performance of the Utilities design could be the small scale of the task. In the experimental scenario, there was only a single robot. This allowed participants to simply always face or listen for the robot while working at the stations and navigating the environment.

P25 [Utilities]: *"It was just easier to look up for the drone when it sounded like it was close/moving."*
P38 [Utilities]: *"I probably used my ears more than the interface. It was helpful and intuitive but it was simpler to listen for the drone."*
P44 [Utilities]: *"I didn't feel the need to use the radar or arrow much because I always faced where the drone was."*

If the space was shared by more than a single robot, it is unlikely that participants would be able to track all robots simultaneously with only sight and sound. We speculate that Utilities design might scale well in this case, providing unobtrusive support for tracking

all proximal robots potentially even better than some of our other designs. However, for the single robot in this experiment, Navpoints, Arrow, and Gaze all performed significantly better than the baseline in terms of decreasing inefficiencies. Participants often noted that explicit visualizations of robot movements made the task easier:

P13 [NavPoints]: *"I found the augmented reality helpful in being able to predict where the robot would move and it made me feel more comfortable with completing the task alongside the robot."*
P33 [NavPoints]: *"It was very helpful while completing the task. I could see clearly where the robot was going to go."*
P17 [Arrow]: "*Once I saw the next two stations the robot was going to, it was easy to work alongside the machine and focus on my work."*
P18 [Gaze]: *"The eyeball design was intuitive and straightforward. It made predicting the drone's movements possible, which made it easier to complete the task."*

Although each of these designs improved objective task performance, only the NavPoints design was consistently rated highly across all subjective scales. Many participants noted that explicitly encoding the robot's arrivals and departures was helpful:

P47 [NavPoints]: *"The circles that indicated how long the robot was going to stay at each station were very useful."*
P50 [NavPoints]: *"The lines and station timers were very helpful."*

However, one aspect of the NavPoints model that was not well-received by participants was the movement speed indicators. Users deemed them to be largely ineffective:

P13 [NavPoints]: *"It was easy to tell where the robot was heading and approximately how long it would be there. It was more difficult to determine how fast the robot would move from station to station."*
P50 [NavPoints]: *". . . the dots showing speed were not as helpful."*

Although the Arrow design provides cues that allow users to infer when the robot would leave a location, participants would have preferred this information be communicated more explicitly. Using the Arrow design, if users wished to know when the robot would leave a station, they had to watch the tip of the arrow indicator depart the station and keep a mental count of how many seconds had passed to know when the robot would also leave that location. This proved to be a mental burden for users:

P29 [Arrow]: *"The arrows gave clear indication about where the robot would fly and how quickly it would leave. At times, I wished for clearer indication of when it would be moving once more."*
P32 [Arrow]: *"Intuitive but needed improvement/more detail about robot movement, such as an indicator of speed and movement pauses."*

The fading opaque/transparent lens in the Gaze design explicitly communicated this information showing users when the robot would leave a location, but it only provided a cue regarding the robot's immediate next destination. Users found this information too limited:

P07 [Gaze]: *"It was tough to plan two steps ahead of the robot but at least I always knew where its next step was going to be."*
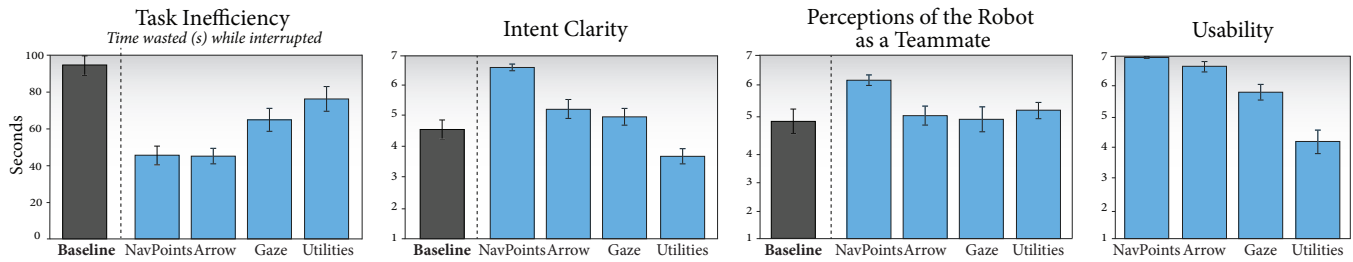
Figure 4: Objective results show that the NavPoints, Arrows, and Gaze designs improved task performance by decreasing inefficiencies and wasted time. Subjective results reveal that NavPoints outperformed other designs in terms of user preferences and perceptions of the robot.

Additionally, one participant had a concern with the appearance of the virtual eye, highlighting the need for careful consideration when altering robot appearance with AR:

P54 [Gaze]: *"I had a pretty good idea of what direction the robot was going to go, but the design made me very nervous (giant floating eyeballs bring killer robots to mind). This probably caused me to be far more hesitant to move towards stations."*

6.1 Limitations and Future Work

While our approach highlights the potential utility of using ARHMD interfaces to convey robot intent, it is not without practical and theoretical limitations. Our current experimental framework relies on a motion tracking system for precise robot localization and navigation, meaning we must translate designs into the motion tracking coordinate space. Future work might merge our approach with developments in robotic perception and simultaneous localization and mapping (SLAM) to remove this constraint, with an ARHMD receiving information directly from the robot regarding its localized position. Other practical limitations may arise from our use of the HoloLens apparatus. While the HoloLens offers significant benefits over prior ARHMDs and may serve as a model for future consumer-grade devices, participants found the headset to be generally uncomfortable with a narrow field of view (FOV).

In addition, the generalizability of our results may be limited due to choices in our experimental design, task, and measures. For instance, our choice of a between-subjects experimental design carries with it inherent trade-offs compared with other approaches, such as a within-subjects design. While a within design would have increased our statistical power, we were concerned it may have introduced detrimental transfer effects, with repeated trials increasing participants' familiarity and comfort with the task and robot while leading to them learn robot behaviors (e.g., orientation signaling movement). Also, several practical concerns made a between-subjects design more feasible (e.g., reducing participant fatigue, robot battery life issues, etc.). Future work might increase participant sample size and/or explore a within-participants design, or employ a different type of task. Our task design had an emphasis on participant efficiency, rather than pure productivity. We observed in our pilot tests that the task had highly variable completion rates, owing to varying degrees of participant manual dexterity; therefore, we did not analyze the total the number of bead strings completed. Future work might examine pure productivity or measure objective trade-offs between productivity, efficiency, and accuracy. For our subjective measures,

we chose to construct custom scales rather than utilize established usability metrics (e.g., the SUS). This choice was motivated by our concern that prior metrics, which have been constructed based on user interaction with traditional computer systems (e.g., GUIs), may not transfer well to novel interactions with flying robotics and AR. Although our scales may be more applicable to the technologies used in this study, more work is needed to compare them with established metrics.

Finally, our designs represent only a subset of all possible design choices using ARHMDs. Our goal is not to claim our designs as optimal, but rather to show the value and potential of the AR-HRI design space and encourage further explorations. Although our designs demonstrated objective and subjective improvements, future work might further explore the concepts we have proposed, especially their generalizability in more chaotic, cluttered, or separated spaces. Alternatively, follow-up studies might explore the scalability of our designs in supporting larger team interactions, with multiple robots and/or multiple people. Future work could also explore gender effects for previous studies have found females to be more responsive to non-verbal cues, including body language in robots [33, 41].

7 CONCLUSIONS

In this work, we explore the design of AR interfaces for conveying robot motion intentions. We propose a framework for structuring the design of AR interfaces that support HRI and use this framework to develop four ARHMD interfaces, which provide various forms of visual feedback for communicating motion intent. We conducted a 60-participant user study comparing these designs against a baseline lacking AR feedback. We found that several of our designs significantly improved user understandings of robot intent and increased objective task efficiency. We also found several trade-offs in our designs, with overall user preferences for more explicit information, especially regarding flight timings. Although wearing an ARHMD may not always be desired or feasible in all scenarios, our research may help inform future studies by demonstrating the promise of AR technologies for mediating human-robot interactions and showcasing the design of novel interface techniques that provide intuitive, visual cues to help users better understand robot intent and as a result improve task efficiency and robot usability.

8 ACKNOWLEDGMENT

REFERENCES

[1] John R Aiello. 1987. Human Spatial Behavior. *Handbook of Environmental Psychology* 1, 1987 (1987), 389–504.

[2] Rasmus S Andersen, Ole Madsen, Thomas B Moeslund, and Heni Ben Amor. 2016. Projecting Robot Intentions into Human Environments. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'16)*.

[3] Asfour, Tamim and Dillmann, Rüdiger. 2003. Human-like Motion of a Humanoid Robot Arm Based on a Closed-Form Solution of the Inverse Kinematics Problem. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'03)*, Vol. 2. 1407–1412.

[4] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47.

[5] Kim Baraka, Stephanie Rosenthal, and Manuela Veloso. 2016. Enhancing Human Understanding of a Mobile Robot's State and Actions using Expressive Lights. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'16)*. 652–657.

[6] John Travis Butler and Arvin Agah. 2001. Psychological Effects of Behavior Patterns of a Mobile Personal Robot. *Autonomous Robots* 10, 2 (2001), 185–202.

[7] 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 *European Conference on Mobile Robots (ECMR'15)*. 1–6.

[8] Mike Daily, Youngkwan Cho, Kevin Martin, and Dave Payton. 2003. World Embedded Interfaces for Human-Robot Interaction. In *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS'03)*. IEEE.

[9] Catherine Diaz, Michael Walker, Danielle Albers Szafir, and Daniel Szafir. 2017. Designing for Depth Perceptions in Augmented Reality. In *Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR'17)*. 111–122.

[10] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and Predictability of Robot Motion. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*. 301–308.

[11] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'93)*. 145–155.

[12] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems* 42, 3–4 (2003).

[13] Foyle, David C and Andre, Anthony D and Hooey, Becky L. 2005. Situation Awareness in an Augmented Reality Cockpit: Design, Viewpoints and Cognitive Glue. In *Proceedings of the 11th International Conference on Human Computer Interaction*. 3–9.

[14] Gielniak, Michael J and Thomaz, Andrea L. 2011. Generating Anticipation in Robot Motion. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'11)*. 449–454.

[15] Matthew Gillen, Joseph Loyall, Kyle Usbeck, Kelly Hanlon, Andrew Scally, Joshua Sterling, Richard Newkirk, and Ralph Kohler. 2012. Beyond Line-of-Sight Information Dissemination for Force Protection. In *IEEE Military Communications Conference (MILCOM'12)*. 1–6.

[16] Scott A Green, Mark Billinghurst, XiaoQi Chen, and J Geoffrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems* 5, 1 (2008).

[17] Scott A Green, J Geoffrey Chase, XiaoQi Chen, and Mark Billinghurst. 2009. Evaluating the Augmented Reality Human-Robot Collaboration System. *International Journal of Intelligent Systems Technologies and Applications* 8, 1-4 (2009).

[18] John Harris and Ehud Sharlin. 2011. Exploring the affect of abstract motion in social human-robot interaction. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'11)*. IEEE, 441–448.

[19] Matthew Hart. 2017. Floating Eyeball Drone is Delightfully Dystopian. (September 2017). http://nerdist.com/floating-eyeball-drone-is-delightfully-dystopian/ [Online; posted 21-March-2016].

[20] Keita Higuchi, Katsuya Fujii, and Jun Rekimoto. 2013. Flying Head: A Head-synchronization Mechanism for Flying Telepresence. In *International Conference on Artificial Reality and Telexistence (ICAT'13)*. 28–34.

[21] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. 2015. Using Gaze Patterns to Predict Task Intent in Collaboration. *Frontiers in Psychology* (2015).

[22] Kentaro Ishii, Shengdong Zhao, Masahiko Inami, Takeo Igarashi, and Michita Imai. 2009. Designing Laser Gesture Interface for Robot Control. *IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT'09)* (2009).

[23] Gary Klein, Paul J Feltovich, Jeffrey M Bradshaw, and David D Woods. 2005. Common Ground and Coordination in Joint Activity. *Organizational Simulation* 53 (2005), 139–184.

[24] Li, Hua and Zhang, Xin and Shi, Guangwei and Qu, Hemeng and Wu, Yanxiong and Zhang, Jianping. 2013. Review and Analysis of Avionic Helmet-Mounted Displays. *Optical Engineering* 52, 11 (2013), 110901–110901.

[25] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems* 77, 12 (1994), 1321–1329.

[26] Paul Milgram, Shumin Zhai, David Drascic, and Julius Grodski. 1993. Applications of Augmented Reality for Human-Robot Communication. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'93)*, Vol. 3. 1467–1472.

[27] Chrystopher L Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, and Rachid Alami. 2005. A Methodological Approach Relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'05)*. 371–377.

[28] Curtis W Nielsen, Michael A Goodrich, and Robert W Ricks. 2007. Ecological Interfaces for Improving Mobile Robot Teleoperation. *IEEE Transactions on Robotics* 23, 5 (2007), 927–941.

[29] Donald A Norman. 1986. Cognitive engineering. *User Centered System Design* 31 (1986), 61.

[30] Shayegan Omidshafiei, Ali-Akbar Agha-Mohammadi, Yu Fan Chen, N Kemal Ure, Jonathan P How, John Vian, and Rajeev Surati. 2015. Mar-cps: Measurable Augmented Reality for Prototyping Cyber-physical Systems. In *AlAA Infotech Aerospace Conference*.

[31] Elena Pacchierotti, Henrik I Christensen, and Patric Jensfelt. 2005. Human-Robot Embodied Interaction in Hallway Settings: A Pilot User Study. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'05)*. 164–171.

[32] Justin Richer and Jill L Drury. 2006. A Video Game-based Framework for Analyzing Human-Robot Interaction: Characterizing Interface Design in Real-time Interactive Multimedia Applications. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'06)*. ACM, 266–273.

[33] J. C. Rosip and J. A. Hall. 2004. Knowledge of Nonverbal Cues, Gender, and Nonverbal Decoding Accuracy. *Journal of Nonverbal Behavior* 28 (2004), 267–286. Issue 4.

[34] Adam W. Ruch. 2010. Videogame Interface: Artefacts and Tropes. In *Videogame Cultures and the Future of Interactive Entertainment Global Conference*.

[35] 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*. 471–478.

[36] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W McOwan, and Ana Paiva. 2011. Automatic Analysis of Affective Postures and Body Motion to Detect Engagement with a Game Companion. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI'11)*. 305–311.

[37] Shin Sato and Shigeyuki Sakane. 2000. A Human-Robot Interface using an Interactive Hand Pointer that Projects a Mark in the Real Work Space. In *IEEE International Conference on Robotics and Automation (ICRA'00)*, Vol. 1.

[38] Jinglin Shen, Jingfu Jin, and Nicholas Gans. 2013. A Multi-view Camera-projector System for Object Detection and Robot-human Feedback. In *IEEE International Conference on Robotics and Automation (ICRA'13)*. 3382–3388.

[39] Ivan E Sutherland. 1965. The Ultimate Display. *Multimedia: From Wagner to Virtual Reality* (1965).

[40] Daniel Szafir. 2015. *Human Interaction with Assistive Free-Flying Robots*. Ph.D. Dissertation. The University of Wisconsin-Madison.

[41] Daniel Szafir and Bilge Mutlu. 2012. Pay attention!: Designing Adaptive Agents that Monitor and Improve User Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. 11–20.

[42] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2014. Communication of Intent in Assistive Free Flyers. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'14)*. 358–365.

[43] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2015. Communicating Directionality in Flying Robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. ACM, 19–26.

[44] Daniel Szafir, Bilge Mutlu, and Terrence Fong. 2017. Designing Planning and Control Interfaces to Support User Collaboration with Flying Robots. *The International Journal of Robotics Research* 36, 5–7, 514–542.

[45] Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing Thought: Improving Robot Readability with Animation Principles. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM, 69–76.

[46] Stefanie Tellex, Ross A Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. 2014. Asking for Help Using Inverse Semantics. In *Robotics: Science and Systems (RSS'14)*, Vol. 2.

[47] J Gregory Trafton, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. 2005. Enabling Effective Human-Robot Interaction using Perspective-taking in Robots. *IEEE Transactions on Systems, Man and Cybernetics* 35, 4 (2005), 460–470.

[48] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating Robotic Navigational Intentions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. 5763–5769.