Mediating Human-Robot Interactions with Virtual, Augmented, and Mixed Reality

Daniel Szafir$^{(\boxtimes)}$ (iD)

Department of Computer Science and ATLAS Institute,
University of Colorado Boulder, Boulder, CO 80309, USA
daniel.szafir@colorado.edu
http://danszafir.com

Abstract. Effective human-robot interaction (HRI) remains a critical barrier to the successful and widespread deployment of robotic technologies. Fundamentally, HRI problems represent breakdowns in *communication*, where poor information exchange between people and robots leads to disfluencies, incorrect mental models, poorly calibrated trust, inadequate situational awareness, etc. In this paper, I argue that the emergence of a new generation of virtual, augmented, and mixed reality (VAMR) technologies is creating an exciting new design space in which to address these problems in communication. To support this argument, I present the results of three experiments demonstrating the value in using modern VAMR technologies to mediate human-robot interactions and discuss various challenges and opportunities in this emerging space at the intersection of several communities, including robotics, VR, graphics, and human-computer interaction.

Keywords: Human-robot interaction (HRI) · Robotics ·
Virtual reality (VR) · Augmented reality (AR) · Mixed reality (MR) ·
VAMR · 3D user interfaces (3DUI)

1 Overview

Robots hold significant promise in benefitting society by supporting human activities across a variety of critical domains, including manufacturing, construction, healthcare, and space exploration. However, in practice, robot deployments remain quite limited because *robots are extremely difficult for people to work with.* A primary source of these difficulties is that humans and robots do not communicate well; people often find robots incomprehensible and have difficulties understanding what a robot can or will do, while robots lack computational models for reasoning about complex human behaviors.

At their core, these issues (aside from purely technical challenges such as limited battery lifetime) are the result of poor information exchange: either the human does not understand the information the robot is conveying, the robot

cannot understand the user's input, or both. These problems are analogous to the *Gulf of Execution* and *Gulf of Evaluation* concepts within the *Human Action Cycle* (Fig. 1), a proposed model describing human interactions with complex systems from the cognitive engineering and human-computer interaction communities [44–46]. At a high level, gulfs of execution arise when users have difficulty translating their high level goals into inputs that a system understands (often because there is a gap between the user's mental model of how a system works and the actual controls/inputs/sequences that the system provides), while gulfs of evaluation occur when users do not understand system feedback and/or have trouble assessing system state.

Translating these concepts to the realm of human-robot interaction, an example of a gulf of execution would be a user attempting to direct a humanoid robot that lacks speech recognition via voice dialog, while an example of a gulf of evaluation would be a user believing that a robot with visually apparent eyes was actively perceiving and tracking the user and the surrounding environment, even if the robot had no actual cameras or visual system. These gulfs are often readily apparent when considering human interaction with humanoid robots due to inaccurate assumptions users commonly hold regarding the link between robot functionality and morphology (e.g., assuming one can talk to a humanoid robot and that the robot will understand), although they may arise with non-humanoid and non-zoomorphic robots as well [22,36].

Breakdowns in human-robot interactions can arise from gulfs of execution, gulfs of evaluation, or both. As a result, a great deal of research in the HRI community has examined how to

Fig. 1. The Human Action Cycle, a model describing user interaction with complex systems, adapted from [44–46]. This model can also be useful in considering human interactions with robots as breakdowns may occur in both the execution stage (inputting information in an attempt to operate or work with a robot to accomplish some goal) and the evaluation stage (understanding what the robot does and whether that advances the user's goal). Breakdowns in different stages of the cycle will likely require different solutions.

improve human-robot communication. For instance, prior work has examined how to reduce gulfs of execution through the development of computational models that enable robots to interpret a large body of potential human inputs such as gaze (e.g., [1,2,30,31]), gestures (e.g., [53,72]), natural language (e.g., [49,71]), and multimodal cues (e.g., [48,57]), enabling users to interact with robots more naturally and intuitively using methods with which they are already familiar.
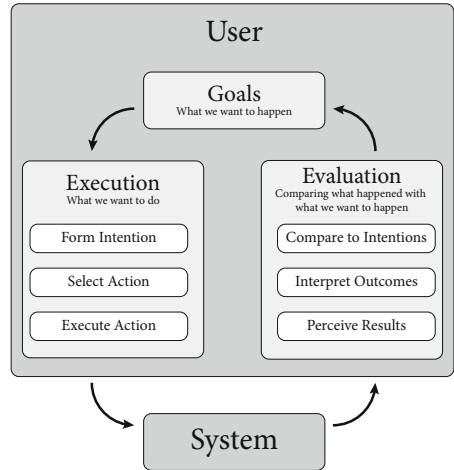
Other work has focused on reducing gulfs of evaluation by conveying robot state and plans via expressive and legible motion [19,61], LED lights [5,62], or other methods (see [15] for a survey). Overall, such research typically must address (at least) two fundamental challenges in identifying (1) the *information content* regarding what information should be communicated and (2) the *information medium* in terms of how the information can be communicated effectively[1].

One of the difficulties in HRI research is that these two challenges of what and how to communicate cannot generally be addressed independently, for the medium chosen will necessarily encode the information content, potentially in a lossy manner or in one that is not easy to decode. For instance, one common medium explored in HRI research is that of gestures, either in research exploring how to interpret human gestures to enhance robot comprehension of user goals or in work exploring how robots might utilize gestures as a means of enhancing interaction fluidity. While such research has shown a variety of benefits for robots using and understanding human gestures, gesture communication itself represents a highly complex information medium with many potential difficulties. For instance, "hash collisions" might occur as many different communicative goals might be mapped to the same or similar gestures (e.g., a closed fist with a single finger raised might indicate a deictic pointing gesture to direct attention to something in the surrounding environment or be used as an iconic gesture indicating the concept of "one"; this challenge has been identified in the conversational agent literature under the maxim "behaviors are not functions" [14]), making it a challenge for robots to interpret human gestures in the absence of other contextual clues. Robot generation of gestures suffers from this same problem and is further complicated by our current lack of standardized robot hardware, meaning desired gestures may be difficult or impossible to implement on a given platform or generalize across platforms. Overall, gesture communication among humans is still not fully understood, thus developing a general framework of gesture modeling and production for robots remains an extremely challenging and open problem.

The fundamental difficulties in communication modeling and understanding described above are not unique to gestural interaction, but are common across any medium used for information exchange (for example, consider the analogous problem of understanding *channel capacity* in information theory). However, different mediums make trade-offs regarding various aspects of communication; for instance verbal communication might be used to communicate larger and more complex ideas than purely gestural communication, but may also require more direct attention be less effective for exchanging immediate or low-level ideas (e.g., a deictic gesture may better direct attention and convey spatial information than a verbal description of where to look). One of the major goals of the HRI field is to bring human-robot interactions to the same level of naturalness and

---

[1] Here "effectively" may take many different meanings depending on context, for instance scenarios involving robots in emergency response might prioritize the speed and accuracy of information exchange, while robots in social settings might prioritize fluidity and naturalness.

effectiveness as human-human interactions (or one day, potentially even surpass them), where interactants commonly switch between mediums and often leverage multiple mediums simultaneously in synchronicity to effectively communicate and identify and repair communicative breakdowns as they occur.

Towards this end, new methods of communication may be highly valuable in enhancing HRI because robots are not limited to only communicating in a manner similar to humans (i.e., using traditional verbal and nonverbal cues). For example, prior work has found that various electronic and computer-mediated methods may be highly beneficial, ranging from the use of enhanced graphical displays to improve robot control (reducing gulfs of execution) [43] to the previously raised example of LED lights that communicate various robot signals (reducing gulfs of evaluation). Recently, the rise of consumer-grade, standardized virtual, augmented, and mixed reality (VAMR) technologies (including the Microsoft HoloLens, Meta 2, Magic Leap, Oculus Rift, HTC Vive, etc.) are creating a promising new medium for information exchange between users and robots. This medium seems particularly well suited to human-robot communication for a variety of reasons. For example, robots often collect 3D spatio-temporal data that may be useful to transmit to users for analysis, which aligns well with one of the primary benefits of the emerging ecosystem of modern VAMR head-mounted displays (HMDs), namely the ability to provide users with 3D virtual imagery with accurate stereoscopic depth cues. In addition, modern HMDs are typically hands-free (and thus may easily integrate with existing solutions for managing human-robot interactions) and can even provide additional communicative channels beyond the visual (e.g., providing 3D spatial audio or microphones for speech recognition).

In this paper, I trace the development of early work merging HRI and VAMR (which, while promising, was often hampered by limitations in underlying VAMR technologies) and highlight more recent work that leverages modern systems. To highlight the potential of VAMR as a communicative medium for mediating HRI, I further detail the results of three laboratory experiments with 126 participants: one experiment examining how VAMR might reduce a gulf of evaluation by presenting users with visualizations of planned robot trajectories, one experiment exploring how VAMR might reduce a gulf of execution by enhancing robot teleoperation, and one experiment examining an integrative VAMR approach towards reducing both gulfs in the design of a new robot supervisory control interface. In each experiment, solutions that utilize VAMR significantly outperform commercially available systems in common use today. I close with a discussion of the current state of VAMR-HRI research and the opportunities and challenges I have observed conducting this research, which sits at the intersection of several communities, including robotics, graphics, and human-computer interaction, that have historically developed largely independent from one-another.

2   Background

The development of VAMR technologies typically trace back to Sutherland's vision of "The Ultimate Display" (itself influenced by Vannevar Bush's

conception of the Memex) [59] and later developments with the Sword of Damocles system [60]. A full review of the development of VAMR technologies and user experience since the 1960's is beyond the scope of this article, although helpful surveys can be found in [4,8,54]. Instead, below I concentrate on tracing early efforts aimed at improving human-robot interaction by integrating VAMR technologies.

2.1   VAMR and HRI

The earliest major work leveraging VAMR for HRI appears to date back to a push in the late 1980's and early 1990's with various work exploring robot teleoperation systems [7,10,26,32,34,56,58,73] that largely originated in the IEEE IROS, ICRA, and SMC communities, although contemporaneously the SIGGRAPH community also noted this as an promising area for future overlapping work [9]. Perhaps the most fully-developed instance of these early systems was the ARGOS interface for augmented reality robot teleoperation [40]. While the ARGOS interface used a stereo monitor, rather than the head-mounted displays in vogue today, the system introduced several design elements for displaying graphical information to improve human-robot communication and introduced concepts such as virtual pointers, tape measures, tethers, landmarks, and object overlays that would influence many subsequent designs (Fig. 2).

Later developments throughout the 1990's introduced several other important concepts, such as the use of virtual reality for both actual robot control and teleoperator training [28,41], the integration of HMDs (including the first use of an HMD to control an aerial robot [66]), projective virtual reality where user "reaches through" a VR system to control a robot that manipulates objects in the real world [24,63], the rise of VAMR applications for robotics in medicine and surgery [11], and continued work on ARGOS and ARGOS-like systems [39,50]. At a high level, major themes appear to be work focused on using VR for simulation or training purposes, VR and/or AR as new forms of information displays (e.g., for data from robot sensors), and generally VAMR-

Fig. 2. The ARGOS system represented one of the earliest robot teleoperation interfaces that leveraged VAMR technology. Here, a "virtual tether" concept is illustrated, which might visualize potential constraints or information about the position and orientation of the manipulator and target. Image reproduced from [40].

based robotic control interfaces. While many of these developments appear initially promising, it is interesting to note that following an initial period of intense early research on HRI and VAMR, later growth throughout the 1990's appears
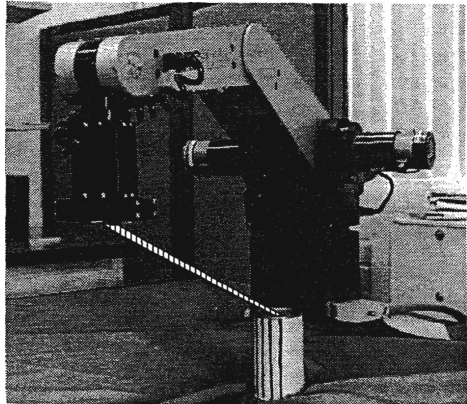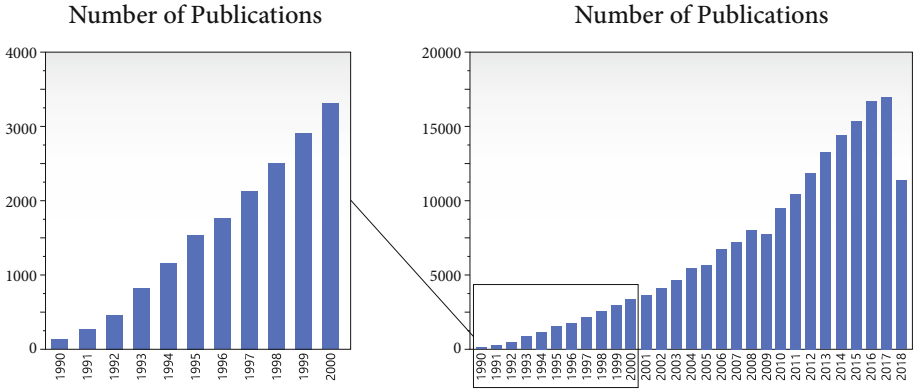
Fig. 3. Publication rates from Google Scholar under the query "'robot' AND 'virtual reality' OR 'augmented reality' OR 'mixed reality'" since 1990 as a rough approximate of research productivity in the VAMR-HRI space. Rates appear to have fairly constant growth between 1990 and 2000 (left), but the full picture (right) indicates we may actually be close to the inflection point for exponential growth and are entering an exciting time for the field. The relatively low numbers for 2018 are likely due to proceedings from 2018 having yet to be archived in Scholar.

to have happened at a relatively stable rate, rather than rapidly expanding (Fig. 3). In addition, efforts to take research developments beyond laboratory environments into commercial/industrial systems appear to have been largely unsuccessful (indeed, even today robot teleoperation interfaces are still typically based on standard 2D displays rather than leveraging VAMR).

In general, I speculate that several key challenges towards VAMR-HRI research may have inhibited the initial growth of the area. First, the lack of any standardized VAMR or robotics hardware at this time introduced an incredibly high barrier towards conducting VAMR-HRI research as essentially all of the research of this time period required laboratories to have the expertise to build both robotics and VAMR equipment, two already highly specialized areas. Moreover, hardware specialization may have introduced difficulties generalizing findings across systems or extending prior work. Second, the general failure of VAMR technologies in the commercial entertainment market of the early 1990's may have soured both industry and academic researchers on conducting further explorations of such systems. This issue may have been exacerbated by the general lack of formal system evaluations in the research developments at this time (i.e., user studies for VAMR-HRI systems of this time period are exceedingly rare, with research papers written largely as system implementations), making it difficult to quantify the value of developments such as those produced in the ARGOS interface.[2] Finally, rather than developing in a cohesive manner,

---

[2] This issue was not unique to VAMR-HRI, as formal evaluation methods have taken time to integrate into many research communities even though they inherently deal with some form of human interaction, including robotics more generally.

research in VAMR-HRI appears to have fragmented throughout the 1990's, with some research extending into more application-driven areas (e.g., robotic surgery or manufacturing), some work examining specific aspects of graphics, visualization, or other VAMR-related areas such as haptics, and other work examining aspects closer to traditional robotics. One critical challenge may have been a lack of a centralized research community and venue for such work, which was also an issue for HRI research more broadly as it had yet to consolidate into its own distinct field (a time that may be demarcated by the first ACM/IEEE HRI conference in 2006). Indeed, by 1999 there was already concern about how "to a large extent the robotics and the newer virtual reality (VR) research communities have been working in isolation" even while there were already clear, promising ideas for their integration [12].

Unfortunately, while certain venues for community consolidation have arisen (e.g., HCII VAMR for both the VAMR and HCI communities), VAMR-HRI work continued in a largely fragmented manner throughout the 2000's, with work scattered across the traditional VAMR communities (IEEE VR, ISMAR, 3DUI, etc.), robotics communities (ICRA and IROS, and eventually HRI and RSS, etc.), and HCI communities (ACM SIGCHI, UIST, etc.) as well as relevant journals (including domain-focused venues, such as for surgical robotics). [25] provides a review of major developments in the early 2000's and a vision for augmented reality HRI as HRI research increasingly focused on various aspects of social interaction and human-robot collaboration. In addition, MiRAs (mixed reality agents) and AuRAs (augmented reality agents) represent a major relevant development during this period, in which robots may interact with or be augmented by virtual agents (e.g., a robot might be shown to a user as driven by a virtual character or have its planned path rendered in physical space) [13, 29].

Overall, while we see an exciting trend in the production of VAMR-HRI papers (in Fig. 3, examining publication rates from 1990 to the present reveals what appears to be the start of an exponential curve, rather than linear growth) and many of the technical limitations in conducting earlier research have been reduced (e.g., due to the increasing prevalence of common/commercially available robot and VAMR platforms), the lack of a centralized research community remains a critical challenge. Towards addressing this issue, the first International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAMR-HRI)[3] was held in 2018 in conjunction with the IEEE/ACM HRI conference, with a followup workshop to occur in March 2019, but it remains to be seen whether this (or other efforts) will help the community converge. Although research fragmentation remains an issue, the overall trajectory for VAMR-HRI research appears very promising, and it is my hope that we are now truly poised to make good on the exciting initial research from the early 1990's. Below, I describe some of my own recent work aligned towards these ends.

---

[3] See http://vam-hri.xyz/.

3   Case Studies

To further build the case for how VAMR can support HRI and how the time is now ripe for a convergence of research that takes advantage of modern hardware in doing so, below I detail some of my own recent work examining the utility of modern VAMR technologies for mediating human-robot interactions. These studies, more fully described in [27,67,68], each examine an aspect of how VAMR might specifically support the Human Action Cycle within the context of a longstanding HRI problem. The first study explores how VAMR can provide visual information to help bridge a gulf of evaluation for users presented with the *motion inference* problem, where people fail to understand when, where, and how a robot teammate will move. The second study explores VAMR in bridging a gulf of execution within the context of *perspective-taking*, i.e., determining what information, and how to convey it, to support robot operators and supervisors in gaining accurate perceptions of the robot and sufficient situational awareness of the robot's working environment to enable precise, and efficient control. The final study harkens back to the early work on VAMR robot teleoperation from the 1990's and combines information gained in the first two experiments towards the design of a comprehensive, modern VAMR teleoperation system that provides new forms of bidirectional information exchange.

3.1   Visualizing Robot Information

One primary challenge towards achieving safe and usable robotic systems is known as the *motion inference* problem, a gulf of evaluation that arises as humans encounter difficulties understanding when, where, and how a robot teammate will move. A great deal of prior HRI work has examined methods for mediating this issue, such as by having robots use human-inspired social cues (e.g., gaze, gestures, etc.) to communicate their intentions [22,42,51], altering robot trajectories to be more legible [19] or expressive [61], or using various other means such as light or auditory indicators [3,6,16,33,47,52,55,62,64,69]. 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, some robots may not be able to reproduce human-based cues due to their morphology, while altering robot motions for legibility or expressiveness may not always be possible in dynamic or cluttered environments, and auditory indicators 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). Instead, we explored VAMR (specifically augmented reality) as an alternative design space for resolving motion inference.

A Design Framework. We began our research process with an analysis how modern VAMR technologies, in particular HMDs that have the advantage of being hands-free, might mediate HRI. Synthesizing information from past VAMR work, including mixed-reality projection systems, VR/AR entertainment applications, and augmented virtuality educational software, we developed a high-level framework for considering how augmented reality HMDs (ARHMDs) might enhance human-robot interactions. 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.*

Briefly, in the first paradigm, virtual imagery is represented as new cues directly embedded into the context of a shared work area using an *environment-as-canvas* metaphor. In the second paradigm, 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, or be used to add full-fledged virtual avatars to physical robots along the MiRA (Mixed Reality Agent) approach [20,29]. In the third 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 [21] and heads-up display technologies used for pilots [23,37]. This third *interface-as-canvas* metaphor may uniquely supply egocentric cues, either directly in front of the user's view or in their periphery, compared to the exocentric feedback provided by augmenting the environment or robot [65]. Overall, we found this augmenting environment/robot/interface framework helpful in surveying the landscape of possible AR interfaces to categorize broad design concepts and in providing us with a structure for reasoning about requirements, benefits, and trade-offs in integrating AR (and possibly, more broadly VAMR) with HRI.

Design Prototypes. We used the design framework described above to develop several reference designs for AR visualizations that might convey robot intent and thus address the motion inference problem. While we prototyped a large number of visualizations, we ultimately ended up evaluating four main designs: *NavPoints* (augmenting environment), *Arrows* (augmenting environment), *Gaze* (augmenting robot), and *Utilities* (augmenting UI). These designs, which 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, can be seen in Fig. 4. At a high level, the NavPoints shows the robot's planned path as 3D waypoints with timers indicating arrival and departure times, the Arrows design provides an arrow

showing the immediate future motion of the robot, the Gaze design provides virtual imagery that alters the robot's morphology such that the robot can make use of gaze behaviors to indicate planned motion in a similar manner as humans, and the Utilities design provides the user with a minimap showing where the robot is in relation to them and gives off-screen indicators if the robot is not currently in view. For more details on each design (including parameters needed for replication), please see [67].

Evaluation. We conducted a 5×1 between-participants experiment to evaluate how our VAMR designs might improve user motion inference when interacting with an autonomous 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 the four designs described above). In the baseline condition, participants still wore an ARHMD (in this study, we used the Microsoft HoloLens), but did not see any virtual imagery. Instead, participants in this condition were informed that the robot had a distinct "front," which always indicated 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. 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.
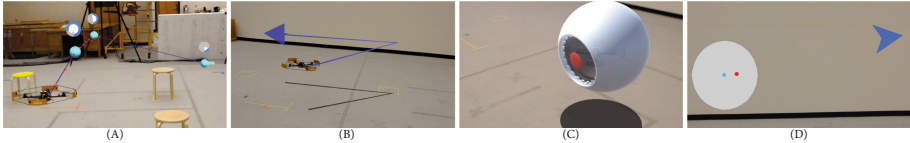

Fig. 4. We explored how augmented reality might address the motion inference problem in HRI by visually conveying robot motion intent. We evaluated four reference prototypes for cuing aerial robot flight motion: (A) NavPoints, (B) Arrows, (C) Gaze, (D) Utilities.

In the experiment, participants had to navigate between several different workstations to collect materials (Fig. 5). These workstations were also used by the robot, who was given priority over the participant, thus participants had to balance their use of the shared resources with being interrupted by the robot. At a high level, the task was set up such that the better participants were at predicting which stations the robot planned to use (i.e., inferring the robot's motion intent), the better they would be able to plan which stations they should use themselves in order to reduce their interruptions and maximize task efficiency.

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 this study. Each experiment lasted approximately 30 min, which included a 60 s tutorial video that provided a brief instruction on the AR feedback participants would receive if they were not in the baseline condition.

Fig. 5. We conducted a laboratory experiment to evaluate the effectiveness of our VAMR designs in improving human-robot interaction. Above, a participant wearing a HoloLens receives AR feedback informing him of the intentions of a nearby robot, helping to bridge a gulf of evaluation.

In this experiment, we found that most of our designs were helpful in improving robot motion inference, enabling participants to more quickly and accurately deduce robot intent in order to plan their own activities more effectively. Analyzing our objective measure of task performance with a one-way Analysis of Variance (ANOVA) using experimental condition (i.e., interface design) as a fixed effect, we found a significant main effect of ARHMD interface design on total time participants 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 participant time lost to interruptions significantly decreased using the NavPoints ($p < .001$), Arrow ($p < .001$), and Gaze ($p = .003$) designs, but not Utilities ($p = .104$). In addition to this objective metric, we also had participants rate several facets regarding their perceptions of the communication of robot movement intent. We found a significant effect of design on perceived communication clarity, $F(4, 55) = 11.04$, $p < .001$, with post-hoc comparisons using Dunnett's test revealing that only the NavPoints design was rated significantly higher than the baseline ($p < .001$). Finally, we compared the designs directly to one another by having participants in all but the baseline condition rate the usability of the displayed virtual imagery for understanding of robot movement intent. We found a significant main effect of design on perceived usability, $F(3, 44) = 25.32$, $p < .001$. 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, $p = .092$. Figure 6 visually summarizes these results.

Overall, we found strong support for the ability of VAMR technology to improve HRI by addressing the motion inference problem. Despite their lack of prior familiarity with VAMR technologies or robots, participants were quickly able to use VAMR feedback displaying robot motion intent and integrate it into their own planning processes, likely due to the intuitive and visual nature of the
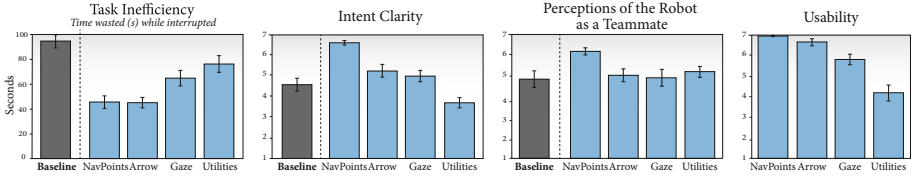
Fig. 6. Objective results show that the NavPoints, Arrows, and Gaze designs improved task performance by decreasing inefficiencies and wasted time. Subjective results reveal that NavPoints outperformed other designs in terms of user preferences and perceptions of the robot.

VAMR designs. We found that our designs that provided more specific information (NavPoints) generally outperformed designs that communicated information in a more implicit manner (Gaze), and that all designs outperformed the Utilities design, which emphasized current robot position relative to the user rather than displaying cues that helped users predict the robot's future destinations. While open questions remain about scalability to scenarios involving larger team interactions with multiple robots and/or multiple people, our results provide strong evidence for the value and potential of the VAMR-HRI design space and showcase the design of novel interface techniques that can provide intuitive, visual cues.

3.2 Information for Control and Supervision

Another critical challenge for HRI is developing interfaces that support effective robot teleoperation and supervision. A substantial body of past research has explored human performance issues in various forms of robotic teleoperation interfaces and mixed teleoperation/supervisory control systems (see [17] for a survey). In particular, prior work has highlighted the issue of *perspective-taking*—the notion that poor perceptions of the robot and its working environment may degrade situational awareness and thus have a detrimental effect on operation effectiveness [17,38].

Mastering perspective taking, where users must rapidly and accurately synthesize information provided directly from the robot (commonly provided via one or more live camera feeds) with an understanding of where the robot is located within the larger context of the environment, is a challenging task, meaning that most robot deployments involving teleoperation still require skilled experts. Current interface designs, particularly for aerial robots (the context we explored in this research), can often exacerbate this problem as live robot camera feeds are typically presented in one of two ways: viewed directly in display glasses or on a traditional screen (e.g., a mobile device, tablet, or laptop computer). While video display glasses may help users achieve an egocentric understanding of what the robot can see, they may degrade overall situational awareness by removing a third-person perspective that can aid in understanding operating context, such as identifying obstacles and other surrounding objects that are not in direct view

of the robot. On the other hand, routing robot camera feeds through traditional displays means that, at any point in time, the operator can only view either the video stream on their display or the robot in physical space. As a result, operators must make constant context switches between monitoring the robot's video feed and monitoring the robot, leading to a divided attention paradigm. To address this issue, we explored how VAMR technologies provide a new medium for designing teleoperation interfaces that can merge viewpoints, enabling teleoperators to monitor the robot in the environment while synchronously monitoring a robot video feed.

Design Prototypes. In exploring perspective-taking, we once again focused on leveraging modern ARHMD technology in the form of the Microsoft HoloLens and utilized the same VAMR-HRI design framework for interfaces that augment the environment/robot/UI described above. Although ARHMD interfaces might provide feedback on many different aspects relevant to robot teleoperation, we focused our design exploration specifically on how to convey information about the robot's camera, as this is typically the most critical information for robots operators.

We developed three primary prototypes, each of which falls within one of the major paradigms in our design framework. We refer to these three design prototypes as *Frustum*, an example of augmenting the environment, *Callout*, an example of augmenting the robot, and *Peripherals*, an example of augmenting the user interface. These designs are each based on prior robot interface designs or other metaphors that may be common to user experiences, adjusted and extended to take advantage of VAMR technology. The Frustum design provides virtual imagery that displays the robot's camera frustum as a series of lines and points, similar to what might be seen emanating from a virtual camera in computer graphics and modeling applications (e.g., Maya, Unity, etc.). The Callout design displays the robot's live camera feed on a panel connected to the top of the robot with an orientation corresponding to the orientation of the camera on the physical robot using a metaphor inspired by speech balloons and thought bubbles. The Peripherals design displays a live robot video feed within a fixed window within the user's view, which was affixed to the periphery of the UI to enable users to monitor the feed while maintaining visual focus on the physical robot in a manner inspired by ambient displays. Each design offers potential tradeoffs in terms of how they support perspective-taking, the total information conveyed, potential scalability across interaction distances, and possibility for user distraction and/or interface overdraw. Figure 7 showcases these interface designs, which are presented in more detail with a discussion of specific design elements in [27].

Evaluation. We conducted a 4×1 between-participants experiment to evaluate how our designs might improve robot teleoperation by improving perspective-taking. The study tasked participants with operating a quadcopter to take several pictures of various targets in a laboratory environment as an analog to aerial
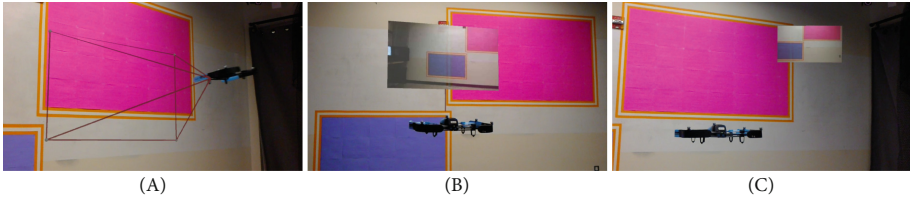
(A)   (B)   (C)

Fig. 7. In this research, we explored how to leverage augmented reality (AR) to improve robot teleoperation. We developed and evaluated 3 design prototypes: (A) the Frustum design augments the environment giving users a clear view of what real-world objects are within the robot's FOV; (B) the Callout design augments the robot like a thought-bubble, attaching a panel with the live video feed above the robot; (C) the Peripheral design provides a window with the live video feed fixed in the user's periphery.

robot inspection and environmental survey tasks. The independent variable in this study corresponded to what type of teleoperation interface the participant used (four levels: Frustum, Callout, and Peripherals designs plus a baseline). In the baseline condition, participants still wore an ARHMD (to control for possible effects of simply wearing a HMD), but did not see any augmented reality imagery. Instead, participants used the Freeflight Pro application[4], the official commercial teleoperation interface supplied by the robot manufacturer. Dependent variables included objective measures of task completion and subjective ratings of operator comfort and confidence.

We recruited a total of 48 participants (28 males, 19 females, 1 self-reported non-binary) from the University of Colorado Boulder campus to participate in our experimental evaluation. Each experiment lasted approximately 30 min, which included two minutes of time spent practicing operating the robot.

The results from our experiment revealed that our VAMR designs outperformed the commercially-available interface across nearly all of our measures. Our objective metrics included task accuracy in terms of the pictures participants took of the experimental targets, task completion time, and number of crashes, each of which we analyzed with a one-way ANOVA with experimental condition as a fixed effect. We found a significant main effect of design on task performance scores for accuracy, $F(3, 44) = 25.01$, $p < .0001$. Comparing designs, Tukey's HSD revealed that the Frustum ($M = 63.2\%$) and Callout ($M = 67.0\%$) interfaces significantly improved inspection performance over the baseline interface ($M = 31.33\%$), with the Peripheral design ($M = 81.1\%$) showing even further benefits by significantly outperforming both Frustum and Callout (all post-hoc results with $p < .0001$). We also found a significant main effect of design on task completion time, $F(3, 44) = 3.83$, $p = .016$. Post-hoc comparisons against the baseline ($M = 239.70$ s) revealed that participants were able to complete the task significantly faster using the Frustum ($M = 140.69$ s), $p = .017$, and Peripherals

---

[4] This popular commercial application can be downloaded from https://itunes.apple.com/us/app/freeflight-pro/id889985763?mt=8.
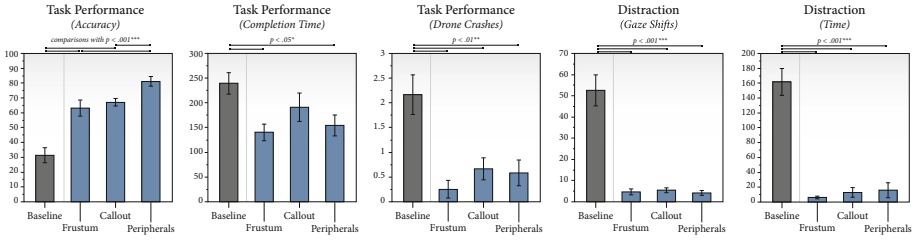
Fig. 8. Objective results show that the augmented reality interface designs improved task performance in terms of accuracy and number of crashes, while minimizing distractions in terms of number of gaze shifts and total time distracted. (*), (**), and (***), denote comparisons with $p < .05$, $p < .01$, and $p < .001$ respectively.

($M = 154.44$ s), $p = .050$, but not the Callout ($M = 191.09$ s), $p = .434$. Examining occurrences when users crashed the robot, we found a significant effect of interface design on operational errors, $F(3, 44) = 9.24$, $p < .001$ with each of our AR designs significantly reducing the number of crashes compared to the baseline (Frustum: $M = .250$, $p < .0001$; Callout: $M = .667$, $p = .003$; Peripherals: $M = .584$, $p = .001$; Baseline: $M = 2.17$).

To better understand these results, we analyzed first- and third-person videos that we recorded of each experiment to look for behavioral patterns. Two coders annotated video data from each interaction based on when participants were able to view the robot and when they were not. Data was divided evenly between coders, with an overlap of 15% of the data coded by both. Inter-rater reliability analysis revealed substantial agreement between raters (Cohen's $\kappa = .92$) [35]. This coding enabled us to calculate *distracted gaze shifts*—the number of times the participant was distracted looking away from the robot during the task; and *distraction time*—the total time spent not looking at the robot. Analyzing this data, we found a significant main effect of interface design on number of distracted gaze shifts, $F(3, 44) = 40.28$, $p < .001$, and on total distraction time, $F(3, 44) = 48.72$, $p < .001$. Post-hoc tests showed that all three VAMR designs significantly decreased both the number and length of distractions compared to the baseline (all comparisons at $p < .0001$), which we take to be evidence that our VAMR prototypes were indeed successful in addressing the perspective-taking issue, thus leading to the performance enhancements found in task accuracy, time, and number of crashes. Figure 8 visualizes these results, while additional analysis of several subjective metrics regarding interface usability, comfort, confidence, etc. that provide further evidence for our conclusions can be found in [27].

Overall, our novel VAMR interface designs that provided users with augmented reality feedback while teleoperating an aerial robot demonstrated significant improvements over a modern interface that is representative of popular designs currently in use. In addition, users rated our designs as more favorable, even though they had a relatively short amount of time with which to practice and may have found the baseline interface, which simply uses a traditional tablet,

more familiar. Once again we believe that these results showcase that interfaces leveraging modern VAMR technologies can be readily integrated with robots to produce highly intuitive user experiences that significantly reduce breakdowns common in human-robot interactions.

3.3   Bidirectional Communication

While each of the studies above showed that VAMR technologies hold promise in helping users bridge a gulf of execution or evaluation, they examined individual aspects of interaction and communicated information in a singular direction (robot-to-human). To more fully examine VAMR-HRI integration within the context of the full Human Action Cycle, we endeavored to design an end-to-end VAMR teleoperation interface. Inspired by the "phantom robot" of [7], our key insight was that VAMR may be used in conjunction with prior work on predictive graphical interfaces such that a teleoperator controls a three-dimensional *virtual robot surrogate*, rather than directly operating the robot itself, providing the user with foresight regarding where the physical robot will end up and how it will get there.

In this system, we provide users with a VAMR robot surrogate—virtual imagery in the form of a "ghost" of the real robot that is embedded within the same operational environment, with accurate stereoscopic depth cues and a matching dynamics model, but that cannot physically interact with the environment (i.e., cannot be damaged or present a hazard to other physical objects or users). This robot avatar serves as a middleman, enabling bidirectional communication of information from human-to-avatar-to-robot and robot-to-human/avatar-to-human. At a theoretical level, we believe that such an interface may help with both gulfs of execution and evaluation by enabling teleoperators to more rapidly iterate through the goal/action/evaluation phases of the Human Action Cycle without the potential of an action that leads to a negative consequence that is realized only after evaluation (e.g., a robot colliding with an obstacle or person). In other words, the system provides users with visuals that let them "test" different inputs and preview results, helping them understand how to pilot the robot to their desired location and providing real-time feedback to understand if course corrections are needed. This system can help reveal mappings between operator input and robot dynamics, information that traditional teleoperation hides in implicit system encodings that users must learn indirectly through experience (e.g., learning the relationship between joystick angle and motor torque).

Design and Evaluation. Robot actions might be tied to actions of a virtual surrogate in a variety of ways. In this research, we explored two main control paradigms: (1) *Realtime Virtual Surrogate (RVS)* operation where the virtual robot responds instantaneously to user input (matching standard forms of teleoperation) while the physical robot, connected to the surrogate with a virtual "fishing line" follows the surrogate after a short delay and (2) *Waypoint Virtual*
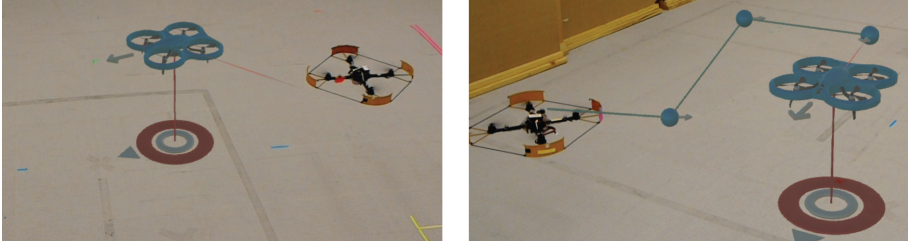
Fig. 9. Two VAMR teleoperation interfaces designed in this work: Left - Realtime Virtual Surrogate (RVS), Right - Waypoint Virtual Surrogate (WVS).

Surrogate (WVS) operation, a delayed form of control that lets the user pilot the surrogate to create a flight plan of various waypoints for the physical robot to traverse, also providing pause/resume and live waypoint editing features that overall leverage AR's ability to place virtual information and objects within the user's environment (see Fig. 9 for visualizations of these interface designs). More specific details on the implementation of each design can be found in [68].

We conducted a 3×1 within-participants experiment to evaluate, relative to a baseline, how our RVS and WVS designs might affect user experiences when teleoperating a collocated aerial robot. In the experiment, participants teleoperated a physical quadcopter to various points of interest (POIs) in a laboratory environment in a task that simulated real-time collection and analysis of environmental data. The independent variable in this study corresponded to what type of teleoperation interface the participant used: a baseine teleoperation system in which the handheld controller input directly controlled the physical robot (i.e., the most common teleoperation system in use today, found in tasks ranging from drone-racing to search-and-rescue), the realtime virtual surrogate design, or the waypoint virtual surrogate system. The same handheld controller and control mapping were utilized across all conditions, although in the RVS and WVS conditions control was rerouted to the surrogate rather than the physical robot. Dependent variables included objective measures of completion time, response time, and interface usage, as well as subjective rankings directly comparing each interface and their perceived multitasking ability, stress, and ease of use.

We recruited a total of 18 participants (11 males, 7 females) from the University of Colorado Boulder campus to take part in this experiment. In our previous two studies, users were largely unfamiliar with either VAMR technologies or robots. While this may be representative of certain target user populations (e.g., users of commercial drones for hobby or entertainment purposes), robot teleoperators in many settings (disaster response, search-and-rescue, etc.) often have a high degree of expertise. As a result, in this experiment we worked to ensure our population sample contained a greater representation of both novices and users experienced at piloting aerial robots. In total, 7 of our participants represented expert users who were recruited from a local "Drone Club," 8 participants

reported moderate familiarity with aerial robots, while 3 participants had little to no experience operating flying robots.

For each participant, the experiment lasted approximately 80 min and consisted of four trials. The first three trials corresponded to the participant using one of the three main interface designs (baseline, RVS, or WVS), with the presented order of interfaces counterbalanced across participants. In the fourth trial, participants were free to use any of the three interfaces and could switch between them at will. Prior to each of the first three trials, participants watched a short 60 s tutorial video that presented the interface design they were going to use, covering both the controls and what the visual feedback looked like (if any). In addition, participants were given two minutes to test each interface before each main trial began, giving them time to become familiar with the controller, augmented reality imagery, and the robot. Participants wore an ARHMD (the Microsoft HoloLens) during all trials as virtual imagery was also used to mark the POIs and show a progress bar corresponding to robot collection of environmental data (even in the baseline condition).

We collected data using a variety of objective and subjective measures to analyze the performance of our two VAMR teleoperation interface designs relative to the baseline of traditional teleoperation. Objective metrics included task completion time and design usage, measured by the percent of total task time that participants used each interface design during the fourth trial in which they were free to switch between interfaces as will (and would presumably use the design(s) they found most helpful). Subjective metrics included the System Usability Scale (SUS), an industry standard ten-item attitude survey for measure perceived usability, several constructed scales to measure aspects of user experience such as stress, and direct rankings to compare each interface in terms of "easy to learn" and "would want to use in the future." We analyzed the objective measures, SUS, and constructed rating scales using a repeated-measures analysis of variance with experimental condition (i.e., interface design) as a fixed effect and condition order included as a covariate to control for potential variance that might arise from ordering effects. Post-hoc tests used Tukey's Honestly Significant Difference (HSD) to control for Type I errors in comparing effectiveness across each interface. Participant rankings of each interface were analyzed with a nonparametric Kruskal-Wallis Test with experimental condition as a fixed effect. Post-hoc comparisons used Dunn's Test for analyzing design sample pairs for stochastic dominance.

Our measures once again revealed the positive benefit that VAMR technologies can have for HRI (see Fig. 10). We found a significant main effect of robot interface design on task completion time, $F(2, 45) = 13.65$, $p < 0.001$, where the RVS ($M = 186.39$ s, $p = .001$) and WVS ($M = 184.39$ s, $p = .001$) designs significantly improved completion time over the baseline interface ($M = 260.11$ s). We also found a significant main effect in regard to design usage during the final trial where participants could switch between designs at will, $F(2, 51) = 34.92$, $p < .001$. Tukey's HSD revealed participants used WVS ($M = 81.94\%$) significantly more than the Virtual Surrogate ($M = 18.06\%$) and Baseline ($M = 0\%$)
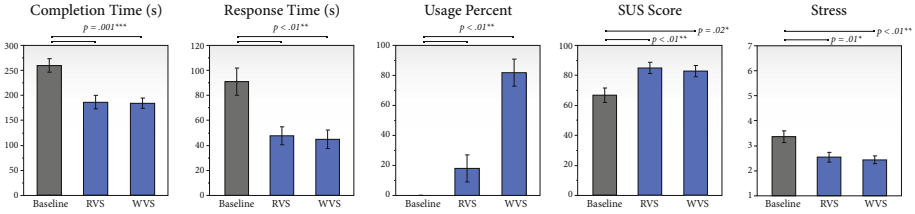
Fig. 10. The RVS and WVS systems showed improvement over the baseline along all objective measures as well as improving subjective user experience (error bars encode standard error).

designs (all comparisons at $p < .001$), with not a single participant ever using the baseline design at any point. We believe this represents extremely strong evidence for the utility of VAMR systems given that even users who were experts in the baseline system (i.e., participants recruited from our local Drone Club) chose to use the VAMR interfaces rather that the control system with which they had prior familiarity. Our subjective results provide additional supporting evidence for the perceived usefulness of the VAMR designs over the baseline (for a full discussion of these results, please see [68]), while highlighting some qualitative differences between RVS and WVS. Although the WVS design was consistently ranked highest in terms of the system users most wanted to use again and by a wide margin the most-used system during the summary evaluation in which users were free to use whichever interface they preferred, it received mixed feedback in open-ended responses where users were asked to comment on each interface. In particular, some users found the WVS system to create too much of a control disconnect between them and the robot. Synthesizing our results and feedback suggests that the RVS system may be most appropriate for hobby use, non-critical tasks, or when users prefer more direct control as it struck a balance between being an enjoyable, responsive, and effective system, while the WVS system may be more useful in professional or multitasking applications where performance trumps user preference.

While the aim of this study was not to design an "optimal" interface, we were nevertheless encouraged by the strong results showing the benefits of integrating VAMR and robotics, building upon prior work in 2D graphical predictive interfaces and the vision of early VAMR teleoperation systems from the 1990's. As mixed, collocated teams of humans and robots become increasingly prevalent in our society, we envision interfaces, such as those evaluated in this and our previous studies, assisting across a variety of human-robot interaction contexts, ranging from robot inspection of equipment and structures, teleoperation on factory floors, and all the way to space exploration where astronauts and/or ground control may be in direct contact or exert supervisory control of various ground and aerial robots.

4   Discussion

In each of these three efforts, we have demonstrated that VAMR designs can lead to significant performance benefits over existing solutions. Much of this work was inspired by ideas first introduced in the early 1990's that can now be more fully realized with modern hardware and validated through empirical experimentation. Such validation is critical for the field to advance beyond technical curiosities and into commercially viable interfaces and software.

In general, we have found that modern VAMR HMDs have several benefits over past systems, including enabling a standardized approach for presenting stereographic imagery, simple or automatic calibration, modern development environments (e.g., Unity and Unreal Engine), onboard (and often built-in) solutions for SLAM, additional built-in sensors and devices such as microphones and speakers, and the capacity for hands-free operation enabling integration with prior systems (e.g., existing teleoperation controllers). However, such systems are not without limitations; for instance, most systems are still limited by field of view (e.g., the HoloLens provides a $30° \times 17.5°$ FOV for virtual imagery), the ability to properly show occlusions with real-world objects, and the ability to be used in bright and/or outdoor environments (although these last two limitations can be mitigated by modern video pass-through technologies, such as the Zed Mini).

One major hurdle we had to face in our research is that there is very limited (or no) support for linking VAMR development libraries with standard robotics development systems (i.e., ROS). In our work, we followed a network communication approach similar to that outlined in [18] to pass data between each system, although new efforts such as ROSBridgeLib[5] and ROS Reality [70] aim to address this issue. However, once a communication layer is established, modern VAMR technologies provide an unprecedented ability for researchers and developers to rapidly prototype HRI designs. For example, in our first study we only ended up evaluating four VAMR designs in our final experiment, but these four designs were downselected through pretesting from an initial candidate set of eight prototypes that each sampled different areas within our design framework.

Another major challenge we have faced is the lack of a theoretical framework to ground VAMR-HRI work. In our research, we have grounded our work in two ways: first, we have leveraged the Human Action Cycle to reason about potential HRI breakdowns that VAMR might address, and second we have developed our model of cues that augment the environment/robot/UI to reason over potential solutions and categorize past work. However, our model is clearly preliminary and may fail to capture nuances across designs and miss other important axis (for instance, it is focused on AR and may be of less use as newer systems increasingly provide the ability to dynamically move along the reality-virtuality continuum). In addition, we have leveraged prior work, where appropriate, to

---

[5] ROSBridgeLib can be downloaded from https://github.com/MathiasCiarlo/ROSBridgeLib.

inspire our design process (e.g., research in graphical predictive interfaces helped motivate our third study). Further work in this area could be greatly aided by a more thorough review of past VAMR-HRI work (at present, no such survey exists) and a cohesive framework to anchor the burgeoning field. In addition, as discussed above, future work would be aided by a greater cohesion across various communities interested in this space, including integrating work from other related fields, such as the graphics and visualization communities, and clear venues to target for publication; at current such work still feels ancillary in either robotics or VAMR venues.

Overall, our work is still limited in many ways. For example, each of our three studies was limited to exploring interactions between a single user and robot in a controlled laboratory environment. As a result, more work is needed to explore scalability to larger, more complex interactions in more realistic conditions (e.g., via field deployments). There are many additional open questions for the research community interested in this space, including developing technical solutions for live, dynamic registration between VAMR technologies and robots, understanding use contexts in which VAMR technologies are and are not appropriate for HRI, improving the development process for VAMR and robotics, integrating additional, related methods of feedback such as haptics, and building information theoretic understandings regarding VAMR, HRI, and communication flow. However, overall we are seeing increasing activity in this space (at least using rough metrics as in Fig. 3) and the time seems ripe for increased innovation, efforts at community-building (e.g., the 2018 and upcoming 2019 VAMR-HRI workshop), and integration across developments from robotics, VAMR, HCI, graphics, and visualization.

5   Conclusion

In this paper, I have briefly traced the origins of work integrating virtual, augmented, and mixed reality technologies with robotics with the goal of improving human-robot interaction. Through three case studies from my own research, I have demonstrated the utility of reviving some of these initial ideas on modern VAMR hardware, leading to objective and experiential improvements over existing commercial systems. In addition, I have introduced the Human Action Cycle as a valuable model borrowed from the cognitive engineering and the HCI communities that can be adapted to reasoning about HRI in general and VAMR-HRI specifically. I have also proposed a preliminary framework for VAMR-HRI developments in terms of providing visual cues that augment the shared environment, the robot(s), or the user interface that, while limited, may serve as a useful starting point in building more formal models for the field. In both my work and VAMR-HRI more broadly, several themes have emerged, including common approaches for leveraging VAMR as a robotics visualization tool, as a control system, or as a training platform as well as common challenges, such as the lack of research cohesion due to fragmentation across several fields. While a formal survey and thematic analysis of the VAMR-HRI research space

over the past thirty years is left for future work, I hope this paper can serve as a rallying cry (or a call to action) for increasing attention to this exciting and growing research area and the need for increased cooperation and interdisciplinary research among the robotics, VAMR, HCI, graphics, and visualization communities.

References

1. Admoni, H., Scassellati, B.: Social eye gaze in human-robot interaction: a review. J. Hum.-Robot Interact. **6**(1), 25–63 (2017)
2. Admoni, H., Srinivasa, S.: Predicting user intent through eye gaze for shared autonomy. In: Proceedings of the AAAI Fall Symposium Series: Shared Autonomy in Research and Practice (AAAI Fall Symposium), pp. 298–303 (2016)
3. Andersen, R.S., Madsen, O., Moeslund, T.B., Amor, H.B.: Projecting robot intentions into human environments. In: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2016), pp. 294–301 (2016)
4. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators Virtual Environ. **6**(4), 355–385 (1997)
5. Baraka, K., Paiva, A., Veloso, M.: Expressive lights for revealing mobile service robot state. Robot 2015: Second Iberian Robotics Conference. AISC, vol. 417, pp. 107–119. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-27146-0_9
6. Baraka, K., Rosenthal, S., Veloso, M.: 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 2016), pp. 652–657 (2016)
7. Bejczy, A.K., Kim, W.S., Venema, S.C.: The phantom robot: predictive displays for teleoperation with time delay. In: IEEE International Conference on Robotics and Automation (ICRA 1990), pp. 546–551 (1990)
8. Billinghurst, M., Clark, A., Lee, G., et al.: A survey of augmented reality. Found. Trends Hum.-Comput. Interact. **8**(2–3), 73–272 (2015)
9. Bricken, W.: Virtual Reality: Directions of Growth Notes from the SIGGRAPH 1990 Panel. Virtual Reality: Directions of Growth, p. 16 (1990)
10. Browse, R., Little, S.: The effectiveness of real-time graphic simulation in telerobotics. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 895–898 (1991)
11. Burdea, G.C.: Virtual reality and robotics in medicine. In: IEEE International Workshop on Robot and Human Communication, pp. 16–25 (1996)
12. Burdea, G.C.: Invited review: the synergy between virtual reality and robotics. IEEE Trans. Rob. Autom. **15**(3), 400–410 (1999)
13. Campbell, A.G., Stafford, J.W., Holz, T., O'Hare, G.M.: Why, When and How to use Augmented Reality Agents (AuRAs). Virtual Reality **18**(2), 139–159 (2014)
14. Cassell, J.: Embodied conversational agents: representation and intelligence in user interfaces. AI Mag. **22**(4), 67 (2001)

15. Cha, E., Kim, Y., Fong, T., Mataric, M.J.: A survey of nonverbal signaling methods for non-humanoid robots. Found. Trends Rob. **6**(4), 211–323 (2018)

16. Chadalavada, R.T., Andreasson, H., Krug, R., Lilienthal, A.J.: 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 2015), pp. 1–6 (2015)

17. Chen, J.Y., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **37**(6), 1231–1245 (2007)

18. Codd-Downey, R., Forooshani, P.M., Speers, A., Wang, H., Jenkin, M.: From ROS to unity: leveraging robot and virtual environment middleware for immersive teleoperation. In: IEEE International Conference on Information and Automation (ICIA 2014), pp. 932–936 (2014)

19. Dragan, A.D., Lee, K.C., Srinivasa, S.S.: Legibility and predictability of robot motion. In: Proceedings of the HRI 2013, pp. 301–308 (2013)

20. Dragone, M., Holz, T., O'Hare, G.M.: Using mixed reality agents as social interfaces for robots. In: IEEE International Symposium on Robot and Human interactive Communication (RO-MAN 2007), pp. 1161–1166 (2007)

21. Feiner, S., MacIntyre, B., Haupt, M., Solomon, E.: Windows on the world: 2D windows for 3D augmented reality. In: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology, pp. 145–155 (1993)

22. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Rob. Auton. Syst. **42**(3–4), 143–166 (2003)

23. Foyle, D.C., Andre, A.D., Hooey, B.L.: Situation awareness in an augmented reality cockpit: design, viewpoints and cognitive glue. In: Proceedings of the 11th International Conference on Human Computer Interaction, pp. 3–9 (2005)

24. Freund, E., Rossmann, J.: Projective virtual reality: bridging the gap between virtual reality and robotics. IEEE Trans. Rob. Autom. **15**(3), 411–422 (1999)

25. Green, S.A., Billinghurst, M., Chen, X., Chase, J.G.: Human-robot collaboration: a literature review and augmented reality approach in design. Int. J. Adv. Rob. Syst. **5**(1), 1 (2008)

26. Grodski, J.J., Milgram, P., Drascic, D.: Real and virtual world stereoscopic displays for teleoperation. In: NATO DRG Seminar: Robotics in the Battlefield (1991)

27. Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018), pp. 78–86 (2018)

28. Hine, B., Hontalas, P., Fong, T., Piguet, L., Nygren, E., Kline, A.: VEVI: a virtual environment teleoperations interface for planetary exploration. SAE Trans. 615–628 (1995)

29. Holz, T., Campbell, A.G., O'Hare, G.M., Stafford, J.W., Martin, A., Dragone, M.: MiRA: mixed reality agents. Int. J. Hum.-Comput. Stud. **69**(4), 251–268 (2011)

30. Huang, C.M., Andrist, S., Sauppé, A., Mutlu, B.: Using gaze patterns to predict task intent in collaboration. Front. Psychol. **6**, 1049 (2015)

31. Huang, C.M., Mutlu, B.: Anticipatory robot control for efficient human-robot collaboration. In: ACM/IEEE International Conference on Human Robot Interaction (HRI 2016), pp. 83–90 (2016)

32. Ince, I., Bryant, K., Brooks, T.: Virtuality and reality: a video/graphics environment for teleoperation. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 1083–1089 (1991)

33. Ishii, K., Zhao, S., Inami, M., Igarashi, T., Imai, M.: Designing laser gesture interface for robot control. In: Gross, T., et al. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 479–492. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03658-3_52

34. Kim, W., Tendick, F., Stark, L.W.: Visual enhancements in pick-and-place tasks: human operators controlling a simulated cylindrical manipulator. IEEE J. Rob. Autom. **3**(5), 418–425 (1987)

35. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**, 159–174 (1977)

36. Li, D., Rau, P.P., Li, Y.: A cross-cultural study: effect of robot appearance and task. Int. J. Soc. Rob. **2**(2), 175–186 (2010)

37. Li, H., Zhang, X., Shi, G., Qu, H., Wu, Y., Zhang, J.: Review and analysis of avionic helmet-mounted displays. Opt. Eng. **52**(11), 110901–110901 (2013)

38. Menchaca-Brandan, M.A., Liu, A.M., Oman, C.M., Natapoff, A.: Influence of perspective-taking and mental rotation abilities in space teleoperation. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2007), pp. 271–278 (2007)

39. Milgram, P., Rastogi, A., Grodski, J.J.: Telerobotic control using augmented reality. In: IEEE International Workshop on Robot and Human Communication (RO-MAN 1995), pp. 21–29 (1995)

40. Milgram, P., Zhai, S., Drascic, D., Grodski, J.: Applications of augmented reality for human-robot communication. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 1993), vol. 3, pp. 1467–1472 (1993)

41. Miner, N.E., Stansfield, S.A.: An interactive virtual reality simulation system for robot control and operator training. In: IEEE International Conference on Robotics and Automation (ICRA 1994), pp. 1428–1435 (1994)

42. Nehaniv, C.L., Dautenhahn, K., Kubacki, J., Haegele, M., Parlitz, C., Alami, R.: A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. In: IEEE RO-MAN 2005, pp. 371–377 (2005)

43. Nielsen, C.W., Goodrich, M.A., Ricks, R.W.: Ecological interfaces for improving mobile robot teleoperation. IEEE Trans. Rob. **23**(5), 927–941 (2007)

44. Norman, D.: Stages and levels in human-machine interaction. Int. J. Man-Mach. Stud. **21**(4), 365–375 (1984)

45. Norman, D.: Cognitive artifacts. Des. Interact.: Psychol. Hum.-Comput. Interface **1**(1), 17–38 (1991)

46. Norman, D.A., Draper, S.W.: User Centered System Design: New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., Mahwah (1986)

47. Omidshafiei, S., et al.: MAR-CPS: measurable augmented reality for prototyping cyber-physical systems. In: AIAA Infotech Aerospace Conference (2015)

48. Perzanowski, D., Schultz, A.C., Adams, W., Marsh, E., Bugajska, M.: Building a multimodal human-robot interface. IEEE Intell. Syst. **16**(1), 16–21 (2001)

49. Prendergast, D., Szafir, D.: Improving object disambiguation from natural language using empirical models. In: Proceedings of the ACM International Conference on Multimodal Interaction (ICMI 2018), pp. 477–485 (2018)

50. Rastogi, A., Milgram, P., Drascic, D., Grodski, J.J.: Telerobotic control with stereoscopic augmented reality. In: Stereoscopic Displays and Virtual Reality Systems III, vol. 2653, pp. 115–123. International Society for Optics and Photonics (1996)

51. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: 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 2011), pp. 305–311 (2011)
52. Sato, S., Sakane, S.: 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 2000), vol. 1, pp. 589–595 (2000)
53. Sauppé, A., Mutlu, B.: Robot deictics: how gesture and context shape referential communication. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, pp. 342–349 (2014)
54. Schuemie, M.J., Van Der Straaten, P., Krijn, M., Van Der Mast, C.A.: Research on presence in virtual reality: a survey. CyberPsychol. Behav. **4**(2), 183–201 (2001)
55. Shen, J., Jin, J., Gans, N.: A multi-view camera-projector system for object detection and robot-human feedback. In: IEEE International Conference on Robotics and Automation (ICRA 2013), pp. 3382–3388 (2013)
56. Spofford, J.R., Garcia, K.D., Gatrell, L.B.: Machine vision augmented displays for teleoperation. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 19–23 (1991)
57. Stiefelhagen, R., Fugen, C., Gieselmann, R., Holzapfel, H., Nickel, K., Waibel, A.: Natural human-robot interaction using speech, head pose and gestures. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 3, pp. 2422–2427 (2004)
58. Stone, R.: Advanced human-system interfaces for telerobotics using virtual reality and telepresence technologies. In: IEEE International Conference on Advanced Robotics, pp. 168–173 (1991)
59. Sutherland, I.E.: The ultimate display. In: Proceedings of IFIP Congress, pp. 506–508 (1965)
60. Sutherland, I.E.: A head-mounted three dimensional display. In: Proceedings of the ACM AFIPS Fall Joint Computer Conference, pp. 757–764 (1968)
61. Szafir, D., Mutlu, B., Fong, T.: Communication of intent in assistive free flyers. In: Proceedings of the HRI 2014, pp. 358–365 (2014)
62. Szafir, D., Mutlu, B., Fong, T.: Communicating directionality in flying robots. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015), pp. 19–26. ACM (2015)
63. Takahashi, T., Sakai, T.: Teaching robot's movement in virtual reality. In: IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS 1991), pp. 1583–1588 (1991)
64. Tellex, S., Knepper, R.A., Li, A., Rus, D., Roy, N.: Asking for help using inverse semantics. In: Robotics: Science and Systems (RSS 2014), vol. 2 (2014)
65. Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E., Schultz, A.C.: Enabling effective human-robot interaction using perspective-taking in robots. IEEE Trans. Syst. Man Cybern. **35**(4), 460–470 (2005)
66. de Vries, S.C., Padmos, P.: Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD. In: Head-Mounted Displays II, vol. 3058, pp. 24–34. International Society for Optics and Photonics (1997)
67. Walker, M., Hedayati, H., Lee, J., Szafir, D.: Communicating robot motion intent with augmented reality. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018), pp. 316–324 (2018)
68. Walker, M., Hedayati, H., Szafir, D.: Robot teleoperation with augmented reality virtual surrogates. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019) (2019, in press)

69. Watanabe, A., Ikeda, T., Morales, Y., Shinozawa, K., Miyashita, T., Hagita, N.: Communicating robotic navigational intentions. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), pp. 5763–5769 (2015)
70. Whitney, D., Rosen, E., Ullman, D., Phillips, E., Tellex, S.: ROS reality: a virtual reality framework using consumer-grade hardware for ROS-enabled robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), pp. 1–9 (2018)
71. Williams, T., Núñez, R.C., Briggs, G., Scheutz, M., Premaratne, K., Murthi, M.N.: A Dempster-Shafer theoretic approach to understanding indirect speech acts. In: Bazzan, A.L.C., Pichara, K. (eds.) IBERAMIA 2014. LNCS (LNAI), vol. 8864, pp. 141–153. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12027-0_12
72. Xiao, Y., Zhang, Z., Beck, A., Yuan, J., Thalmann, D.: Human-robot interaction by understanding upper body gestures. Presence: Teleoperators Virtual Environ. **23**(2), 133–154 (2014)
73. Zhai, S., Milgram, P.: Telerobotic virtual control system. In: Cooperative Intelligent Robotics in Space II, vol. 1612, pp. 311–323. International Society for Optics and Photonics (1992)